JMB

Biophysical Properties of Human Antibody Variable Domains

Stefan Ewert, Thomas Huber, Annemarie Honegger and Andreas Plückthun*

Biochemisches Institut
Universität Zürich
Winterthurerstr. 190
CH-8057 Zürich, Switzerland

There are great demands on the stability, expression yield and resistance to aggregation of antibody fragments. To untangle intrinsic domain effects from domain interactions, we present first a systematic evaluation of the isolated human immunoglobulin variable heavy (V_H) and light (V_L) germline family consensus domains and then a systematic series of V_H–V_L combinations in the scFv format. The constructs were evaluated in terms of their expression behavior, oligomeric state in solution and denaturant-induced unfolding equilibria under non-reducing conditions. The seven V_H and seven V_L domains represent the consensus sequences of the major human germline subclasses, derived from the Human Combinatorial Antibody Library (HuCAL®). The isolated V_H and V_L domains with the highest thermodynamic stability and yield of soluble protein were V_H3 and $V_{\kappa}3$, respectively. Similar measurements on all domain combinations in scFv fragments allowed the scFv fragments to be classified according to thermodynamic stability and *in vivo* folding yield. The scFv fragments containing the variable domain combinations H3κ3, H1bκ3, H5κ3 and H3κ1 show superior properties concerning yield and stability. Domain interactions diminish the intrinsic differences of the domains. ScFv fragments containing V_{λ} domains show high levels of stability, even though V_{λ} domains are surprisingly unstable by themselves. This is due to a strong interaction with the V_H domain and depends on the amino acid sequence of the CDR-L3. On the basis of these analyses and model structures, we suggest possibilities for further improvement of the biophysical properties of individual frameworks and give recommendations for library design.

© 2003 Elsevier Science Ltd. All rights reserved

Keywords: antibody engineering; protein stability; expression; scFv fragment

*Corresponding author

Introduction

Because of their high degree of specificity and broad target range, antibodies have found numer-

Present address: S. Ewert, ESBA Tech AG, Wagistr. 21, CH-8952 Zürich-Schlieren, Switzerland.

Abbreviations used: CDR, complementary-determining region; GdnHCl, guanidine hydrochloride; HuCAL®, Human Combinatorial Antibody Library; IMAC, immobilized metal ion affinity chromatography; SB, super broth; scFv, single-chain antibody fragment consisting of the variable domains of the heavy and of the light chain connected by a peptide linker; V_H, variable domain of the heavy chain of an antibody; V_L, variable domain of the light chain of an antibody.

E-mail address of the corresponding author: plueckthun@biocfebs.unizh.ch

ous applications in basic research, medicine and the biotechnology industry, where they serve as tools to recognize selectively virtually any kind of target molecule. However, despite their versatility, only a subset of antibodies has the biophysical properties ideally suited for such applications. For example, therapeutic or *in vivo* diagnostic antibody fragments require a long serum half-life in human patients to accumulate at the desired target, and they must therefore be resistant to aggregation, precipitation and degradation by proteases.[1,2] Industrial applications often demand antibodies that have a long half-life or can function in organic solvents, surfactants or at high temperatures.[3,4] Applications in functional genomics will require stability against surface denaturation and drying in miniaturized devices. Furthermore, the high

demand for and the increasing number of applications of antibodies require more efficient methods for their high-level production and the availability of molecules with favorable properties for every antigen. Single-chain Fv (scFv) fragments are a particularly popular antibody format.[5–7] The size of these molecules is reduced to the antigen-binding part of the antibody, and they contain the variable domains of the heavy and light chain connected *via* a flexible linker. Most scFv fragments can be obtained easily from recombinant expression in *Escherichia coli* in sufficient amounts.[8] As production yields of these fragments are influenced by their stability as well as their folding efficiency, considerable efforts have been made to identify positions in the variable domains critical for influencing the expression behavior of various antibody molecules.[9–12] The factors influencing the biophysical properties of antibody molecules have been studied mostly with scFv fragments.[13] The overall stability of scFv fragments depends on the intrinsic structural stability of V_H and V_L as well as on the extrinsic stabilization provided by their interaction.[14] For some scFvs, the stabilities of isolated V_H and V_L domains as well as of the whole scFv fragment have been measured and compared.[14–16]

In this study, we present the first fully systematic biophysical characterization of isolated human V_H and V_L domains as well as an investigation of the effect of combinations of V_H and V_L in the scFv format. The analyzed variable antibody domains are family consensus domains taken from the Human Combinatorial Antibody Library (HuCAL®).[17] In this library, a consensus sequence of the framework sequences of the major V_H-families (V_H1, V_H2, V_H3, V_H4, V_H5, and V_H6) and V_L-families ($V_\kappa1$, $V_\kappa2$, $V_\kappa3$, $V_\kappa4$ and $V_\lambda1$, $V_\lambda2$, $V_\lambda3$) was generated. CDR-1 and CDR-2 sequences were taken from the germline gene with the highest usage in a family as calculated from rearranged sequences. The V_H1 family is further divided into V_H1a and V_H1b because of different CDR-H2 conformations. We chose these seven V_H and V_L domains for our study, because they ideally represent the structural repertoire of human V_H and V_L frameworks, they were engineered to contain identical CDR-3s and their individual performance in recombinant libraries has been studied extensively.[17,18]

Human V_H domains occur in nature only in complex with V_L. Most isolated V_H domains have a pronounced tendency to aggregate, due primarily to the large hydrophobic surface area exposed in the absence of V_L.[16,19,20] This makes their expression and purification laborious, and the study of their biophysical properties in isolation almost impossible. The experiments reported here became possible, nevertheless, because we recently discovered by a novel selection strategy a 17 amino acid residue long CDR-H3, which increases the solubility of V_H domains, possibly by partially covering the

hydrophobic interface region to V_L (J. Burmester *et al.*, unpublished results).

In contrast to V_H domains, isolated V_L domains usually show only a small tendency to aggregate,[20,21] apart from some Bence–Jones proteins, of which the fibril-forming ones may be identifiable by "structural risk factors".[22] However, the CDR-L3 is again of critical importance, because it differs from V_κ to V_λ domains. The majority of the CDR-L3 (70%) of human V_κ-domains has a *cis*-proline residue at position 136 (the nomenclature is described by Honegger & Plückthun[23]), which defines a rigid Ω-loop conformation, while a *cis*-proline residue is observed only rarely in V_λ domains (4%). The CDR-L3 Ω loop conformation of V_κ is not compatible with the CDR-L1 conformation usually observed in V_λ domains, as the combination of a rigid V_κ CDR-L3 with a V_λ CDR-L1 leads to steric clashes, destabilizing the chimeric domains severely. Taking this into account, we used two different CDR-L3s for V_κ and V_λ domains.

The seven human consensus V_H domains with the solubility-enhancing CDR-H3, the seven human consensus V_L domains with a κ-like CDR-L3 for κ domains and λ-like CDR-L3 for the λ domains, were expressed and purified, and a classification of human V_H and V_L frameworks is given, on the basis of their yield of soluble periplasmic protein, examination of their oligomeric state in solution, denaturant-induced unfolding and refolding equilibria under non-reducing conditions. We further examined scFv fragments with systematic combinations of the most stable V_H domain (V_H3) with each V_L domain and with the most stable V_L domain ($V_\kappa3$) with each V_H domain.

With this study it is now possible to make predictions about scFv properties such as yield, oligomeric state and thermodynamic stability, just by identifying the V_H and V_L family. Most important, however, the detailed analysis of the individual properties of the consensus variable domains and comparisons of sequence alignments and model structures enables a further improvement of properties of any immunoglobulin variable domain and of libraries on the basis of incorporating any of the frameworks, to maintain the full diversity of antibodies. It gives some insight into interacting groups of residues and their co-evolution seen in the variable domain families.

Results

Expression and protein purification of V_H fragments

The seven HuCAL® consensus V_H domains representing the major framework subclasses were expressed with the same CDR-H3 to enable the comparison of their biophysical properties. We first investigated V_H domains with the CDR-H3

Table 1. Summary of biophysical characterization of isolated V_H and V_L domains

	Human family	CDR-3	Soluble yield (mg ml^{-1} OD$_{550}$ = 10)	Oligomeric state	Midpoint [GdnHCl] (M)	ΔG_{N-U} (kJ mol^{-1})	m (kJ M^{-1} mol^{-1})
V_H	1a	Long[a]	1.0	M[b]	1.5	13.7	10.1
	1b	Long	1.2	M	2.1	26.0	12.7
	2	Long	Refold.[c]	ND[d]	1.6	ND	ND
	3	Long	2.4	M	3.0	52.7	17.6
	3[e]	Short[f]	2.1	ND	2.7	39.7	14.6
	4	Long	Refold.	ND	1.8	ND	ND
	5	Long	Refold.	M	2.2	16.5	7.0
	6	Long	Refold.	ND	0.8	ND	ND
V_L	κ1	κ-like[g]	4.5	M	2.1	29.0	14.1
	κ2	κ-like	14.2	M	1.5	24.8	16.1
	κ3	κ-like	17.1	M	2.3	34.5	14.8
	κ4	κ-like	9.6	D, M[h]	1.5	ND	ND
	λ1	λ-like[i]	0.3	M	2.1	23.7	11.1
	λ2	λ-like	1.9	M	1.0	16.0	16.2
	λ3	λ-like	0.8	D, M[h]	0.9	15.1	15.9

[a] Long CDR-H3, sequence: YNHEADMLIRNWLYSDV.
[b] Monomer (M) in 50 mM sodium phosphate (pH 7.0) and 500 mM NaCl, in case of V_H1a with 0.9 M GdnHCl.
[c] No soluble protein obtained, purification *via* refolding of inclusion bodies.
[d] Not determined.
[e] Data from Ewert *et al.*[25]
[f] Short CDR-H3, sequence: WGGDGFYAMDY.
[g] κ-Like CDR-L3, sequence: QQHYTTPPT.
[h] Dimer (D) and monomer (M) equilibrium.
[i] λ-Like CDR-L3, sequence: QSYDSSLSGVV.

(WGGDGFYAMDY) from the antibody hu4D5-8 (PDB entry 1FVC[24]), but only the V_H3 domain could be purified from the soluble fraction of the bacterial lysate,[25] while the other V_H domains by themselves were insoluble and gave small inclusion body pellets (data not shown). This was not surprising, as many if not most isolated V_H domains have been found to be insoluble upon periplasmic expression,[16,19,20,26] since they contain an exposed large hydrophobic interface that is usually covered by V_L. However, three isolated V_H domains from the HuCAL® (with framework classes V_H1a, V_H1b, and V_H3) were obtained recently from a metabolic selection experiment (J. Burmester *et al.*, unpublished results). These could be expressed in the periplasm of *E. coli* and purified from the soluble fraction of the cell extracts. The main feature of the selected V_H domains is the length of the CDR-H3, as all three selected and soluble V_H fragments contain a CDR-H3 that is longer than 15 residues. This long CDR-H3 may partially cover the hydrophobic interface of V_H, thereby preventing aggregation. After introducing the long CDR-H3 from one of the selected V_H3 domains (YNHEADMLIR-NWLYSDV), V_H1a, V_H1b and V_H3 could be expressed in soluble form in the periplasm of *E. coli* and purified from the soluble fraction of the cell extracts with a yield of 2.4 mg l^{-1} in the case of V_H3 and of 1.0 mg l^{-1} and 1.2 mg l^{-1} in the case of V_H1a and V_H1b, respectively (Table 1).

In contrast, V_H2, V_H4, V_H5 and V_H6 with the long CDR-H3 were still insoluble in the *E. coli* periplasm. Protein production was induced with a final concentration of IPTG of 0.05 mM, followed by an incubation time of 15 hours. The low concen-

tration of IPTG caused a slower protein expression that was necessary, because standard concentrations of IPTG between 0.1 mM and 1 mM resulted in cell lysis after two to three hours (data not shown). The domains were purified from the insoluble fraction with immobilized metal ion affinity chromatography (IMAC) under denaturing conditions, and the eluted fractions were subjected to *in vitro* refolding. Approximately 1 mg of soluble, refolded V_H5 domain could be obtained from one liter of *E. coli* culture using an oxidizing glutathione redox shuffle, while V_H2, V_H4 and V_H6 gave disulfide-linked multimers under the same conditions (data not shown). The even-numbered V_H domains could be refolded only by using a redox shuffle with an excess of reduced glutathione followed by extensive dialysis against 50 mM sodium phosphate (pH 7.0), 100 mM NaCl and slow oxidation by dissolved air-oxygen. Finally, about 0.2 mg of soluble, refolded protein could be obtained from one liter of *E. coli* culture. V_H1a, V_H1b, V_H3 and V_H5 remained in solution at 4 °C and no degradation was observed. In contrast, V_H2, V_H4 and V_H6 have a great tendency to aggregate during storage at 4 °C. Therefore, all subsequent experiments were performed with freshly purified proteins.

### Analytical gel-filtration of V_H fragments

To analyze the oligomeric state of the purified V_H domains in solution, analytical gel-filtration experiments were performed. V_H1b, V_H3, and V_H5 elute at the expected size of a monomer (Figure 1(a) with V_H3 as an example for monomeric V_H domains). V_H1a elutes under native conditions in
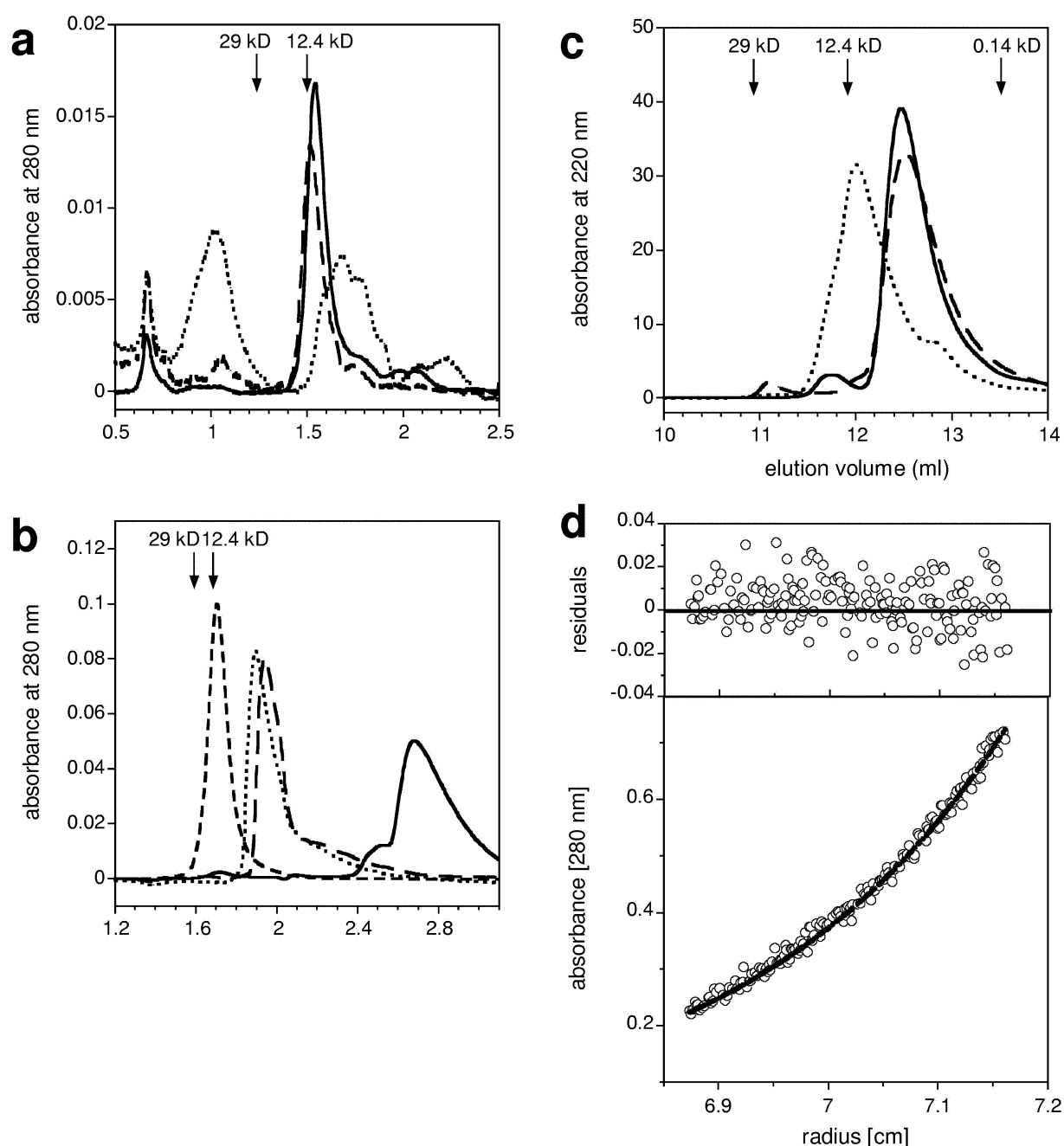
Figure 1. Determination of apparent molecular mass of isolated V_H and V_L domains. Gel-filtration runs were performed in 50 mM sodium phosphate (pH 7.0), 500 mM NaCl of (a) isolated human consensus V_H domains (5 μM) on a Superdex-75 column with V_H3 (continuous line) and V_H1a (dotted line) and V_H1a in the presence of 0.9 M GdnHCl (long-dash line); (b) isolated $V_κ$ domains (50 μM) on a Superose-12 column with $V_κ1$ (continuous line), $V_κ2$ (long-dash line), $V_κ3$ (dotted line) and $V_κ4$ (short-dash line); and (c) isolated $V_λ$ domains (5 μM) on a TSK column with $V_λ1$ (continuous line), $V_λ2$ (long-dash line) and $V_λ3$ (dotted line). Arrows indicate elution volumes of molecular mass standards: carbonic anhydrase (29 kDa), and cytochrome *c* (12.4 kDa). (d) Equilibrium sedimentation of $V_κ3$ at 19,000 rpm (29,100 *g*) with a detection wavelength of 280 nm. The continuous line was obtained from fitting of the data to a single species, and a molecular mass of 13,616 Da was calculated. The residuals of the fit are scattered randomly, indicating that the assumption of the monomeric state is valid.

three peaks that could not be assigned. We therefore investigated whether small amounts of denaturant might break up the aggregates. Using an elution buffer containing 0.5 M GdnHCl, the unassigned peaks decrease and a peak at the size of a monomer was found. With 0.9 M GdnHCl,

V_H1a elutes in a single peak corresponding to a monomer (Figure 1(a)) with the elution profile of V_H1a at 0 M and 0.9 M GdnHCl). It should be noted that this concentration of GdnHCl is below the transition in equilibrium unfolding (see below). V_H2, V_H4 and V_H6 did not elute from the
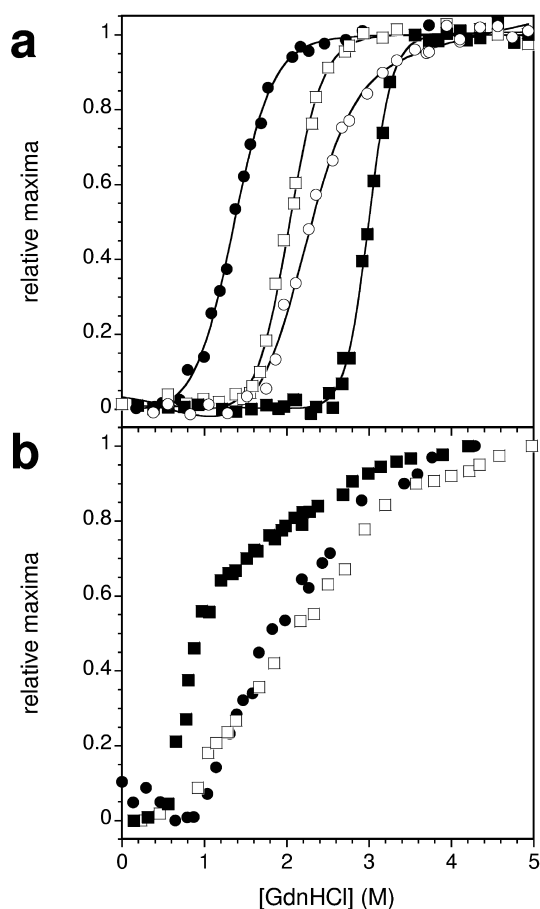
Figure 2. Overlay of GdnHCl denaturation curves of V_H domains (a) V_H1a (filled circles), V_H1b (open squares), V_H3 (filled squares) and V_H5 (open circles). (b) V_H2 (filled circles), V_H4 (open squares) and V_H6 (filled squares). (a) and (b) All unfolding transitions were measured by following the change in emission maximum as a function of denaturant concentration at an excitation wavelength of 280 nm. Relative maxima refers to a scaling in which the lowest emission maximum wavelength is set to 0, the highest to 1. Note that this procedure does not flatten the pre-transition or post-transition baseline.

column under native conditions. Even addition of 1.7 M GdnHCl to the elution buffer did not prevent these domains from sticking to the column. Elution could be achieved only with 1 M NaOH.

Equilibrium transition experiments of V_H fragments

The thermodynamic stability of the seven human consensus V_H domains with the long CDR-H3 was examined by GdnHCl equilibrium denaturation experiments. Unfolding of the V_H domains was monitored by the shift of the fluorescence emission maximum as a function of denaturant concentration. Figure 2(a) shows an overlay of the equilibrium denaturation curves of V_H1a, V_H1b, V_H3, and V_H5. The equilibrium denaturation of these domains under non-reducing conditions is cooperative and reversible (data not shown),

which is consistent with two-state behavior. The V_H1a domain starts to unfold at 0.9 M GdnHCl, where V_H1a is monomeric in solution as indicated by gel-filtration analysis (Figure 1(a)). Therefore, the transition is influenced by only the stability of the monomeric V_H1a domain and is not affected by multimerization equilibria. For the determination of the free energy of unfolding, the pre-transition region of V_H1a, whose actual slope is influenced by the spectral changes caused by dissociation, was assumed to have the same slope and intercept as the V_H1b domain. V_H3 displays the highest change in free energy upon unfolding (ΔG_{N-U}) with 52.7 kJ mol^{-1} and an unfolding cooperativity (m value) of 17.6 kJ mol^{-1} M^{-1}. V_H1b is of intermediate stability with a ΔG_{N-U} of 26.0 kJ mol^{-1} and an m value of 12.7 kJ mol^{-1} M^{-1}. V_H1a and V_H5 are less stable and have ΔG_{N-U} values of 13.7 kJ mol^{-1} and 19.1 kJ mol^{-1}, and m values of 10.1 kJ mol^{-1} M^{-1} and 8.6 kJ mol^{-1} M^{-1}, respectively (Table 1). The range of m values can be compared to that expected for proteins of this size (14–15 kDa), and indicate that at least V_H1a, V_H1b, and V_H3 have the cooperativity expected for a two-state transition.[27] The transition curves of V_H2, V_H4 and V_H6 in Figure 2(b) show poor cooperativity, which indicates that no two-state behavior is followed during GdnHCl equilibrium denaturation. As the monomeric state of these V_H domains could not be ascertained, it is likely that part of this complicated transition involves the dissociation of multimers. The broad transition of V_H2 and V_H4 occurred between 1.0 M and 2.5 M GdnHCl with a midpoint of 1.6 M and 1.8 M GdnHCl, respectively. V_H6 shows a transition between 0.5 M and 1.4 M GdnHCl with a midpoint of 0.8 M. This is the lowest midpoint of the domains examined, which indicates that V_H6 is the least stable human V_H domain.

Expression and protein purification of V_L fragments

The four human consensus V_κ domains (V_κ1, V_κ2, V_κ3 and V_κ4) carrying the κ-like CDR-L3 (sequence: QQHYTTPPT) from the antibody hu4D5-8 (PDB entry 1FVC[24]) were expressed in soluble form in the periplasm of *E. coli*. After purification with IMAC followed by a cation-exchange column the V_κ domains could be obtained in large amounts, ranging from 17.1 mg l^{-1} of bacterial culture normalized to an A_{550} of 10 for V_κ3 to 4.5 mg for V_κ1 (Table 1).

The κ-like CDR-L3 has a *cis*-proline residue at position 136 (the numbering scheme for the variable domain residues is described by Honegger & Plückthun[23]). While 70% of the human V_κ domains have *cis*-Pro in position L136, only 4% of the human V_λ domains show Pro in this position. Therefore, we used a human consensus λ-like CDR-L3 for the V_λ domains (sequence QSYDS-SLSGVV; for details, see Materials and Methods). The three human consensus V_λ domains (V_λ1, V_λ2

and $V_\lambda 3$) were expressed in soluble form in the periplasm of *E. coli*, but the yield after purification with IMAC and anion-exchange column was much less than for the V_κ domains, ranging from 1.9 mg l^{-1} of bacterial culture normalized to an A_{550} of 10 for $V_\lambda 2$ to 0.3 mg l^{-1} for $V_\lambda 1$ (Table 1).

Analytical gel-filtration of V_L fragments

While the monomeric V_H fragments elute at the expected molecular mass around 13 kDa (Figure 1(a)), V_L domains in 50 mM sodium phosphate (pH 7.0), 500 mM NaCl bind to gel-filtration column materials, resulting in an elution profile in which the "affinity separation" of the different oligomeric species on the column material overshadows the gel-filtration effect. In the case of V_κ domains, the best results could be obtained with a Superose-12 column (Figure 1(b)). At a protein concentration of 50 μM, $V_\kappa 3$ and $V_\kappa 2$ elute at an apparent molecular mass of only 2 kDa, $V_\kappa 4$ at 12 kDa and $V_\kappa 1$ elutes as a broad peak at a volume larger than the total volume of the column. Changing the concentration of $V_\kappa 4$ from 5 μM to 50 μM, the peak shifts from an apparent molecular mass of 2–12 kDa, indicating a concentration dependent monomer dimer equilibrium, assuming that V_κ domains eluting at 2 kDa are indeed monomeric and at 12 kDa are dimeric (see below). Addition of 1 M GdnHCl or adjusting the NaCl concentration to 2 M did not alter the elution profile. V_λ domains at concentrations of 5 μM show the weakest unspecific interaction with silica-based TSK columns (Figure 1(c)), and $V_\lambda 1$ and $V_\lambda 2$ elute at a molecular mass of 7 kDa and $V_\lambda 3$ elutes at an apparent molecular mass of 12 kDa.

To interpret these results from analytical gel-filtration, we analyzed the samples by equilibrium ultracentrifugation. We used this method to calibrate the elution volumes of the different columns for V_L domains: $V_\kappa 3$ (shown in Figure 1(d) as an example) and $V_\lambda 2$ give results consistent with a monomer, while $V_\lambda 3$ shows a monomer-dimer equilibrium. Therefore, we can conclude that the V_L domains $V_\kappa 2$, $V_\kappa 3$, $V_\lambda 1$ and $V_\lambda 2$, eluting at an apparent molecular mass at 6 kDa and 2 kDa, respectively, are indeed monomeric and the V_L domains $V_\kappa 4$ and $V_\lambda 3$ eluting at 12 kDa are dimeric. $V_\kappa 1$, which elutes even after the total volume of the column, indicating a strong interaction with the column material, behaves in the ultracentrifugation as a monomer (Table 1).

Equilibrium transition experiments of V_L fragments

Most V_L domains have only one tryptophan residue (the highly conserved Trp L43), which is buried in the core in the native state. Under native conditions, no emission maxima could be determined, because the fluorescence is quenched fully by the disulfide bond Cys L23-Cys L106. During GdnHCl-induced unfolding, the tryptophan resi-
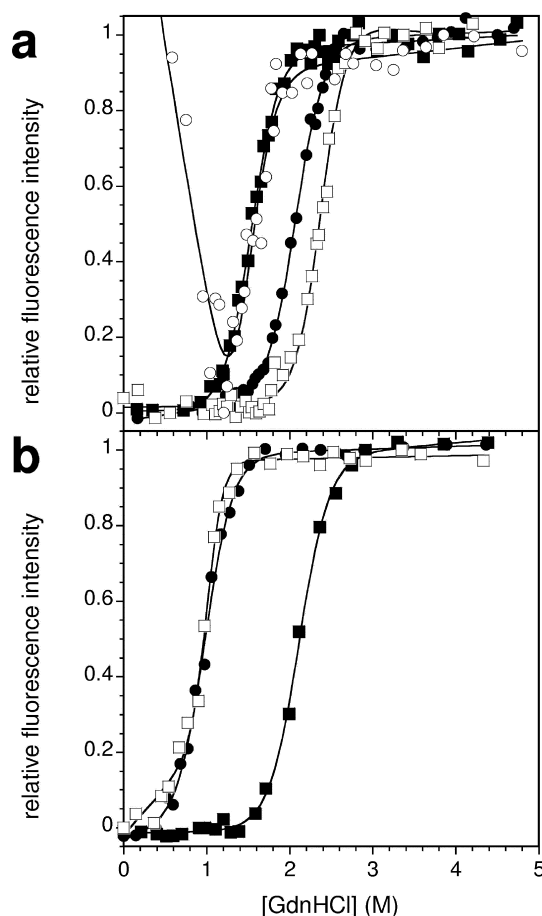
Figure 3. Overlay of GdnHCl denaturation curves of V_L domains (a) V_κ domains with $V_\kappa 1$ (filled circles), $V_\kappa 2$ (filled squares), $V_\kappa 3$ (open squares) and $V_\kappa 4$ (open circles) and (b) V_λ domains with $V_\lambda 1$ (filled squares), $V_\lambda 2$ (filled circles), $V_\lambda 3$ (open squares). (a) and (b) All unfolding transitions were measured by following the change of fluorescence intensity as a function of denaturant concentration at an excitation wavelength of 280 nm.

due becomes solvent-exposed, giving a steep increase in fluorescence intensity. Therefore, the thermodynamic parameters were calculated using the six-parameter fit[28] on the plot of concentration of GdnHCl *versus* fluorescence intensity, giving curves consistent with two-state behavior. All V_L domains show reversible unfolding behavior under non-reducing conditions (data not shown). Figure 3(a) and (b) show relative fluorescence intensity plots against GdnHCl concentration of V_κ and V_λ domains. $V_\kappa 3$ is the most stable V_L domain with a ΔG_{N-U} of 34.5 kJ mol^{-1}, followed by $V_\kappa 1$ with 29.0 kJ mol^{-1} and $V_\kappa 2$ and $V_\lambda 1$ with 24.8 kJ mol^{-1} and 23.7 kJ mol^{-1}, respectively (Table 1). The least stable V_L domains are $V_\lambda 2$ and $V_\lambda 3$ with a ΔG_{N-U} of 16.0 kJ mol^{-1} and 15.1 kJ mol^{-1}. All V_L domains show m values between 11.1 kJ mol^{-1} M^{-1} and 16.2 kJ mol^{-1} M^{-1}, indicating that they have the cooperativity expected for a two-state transition.[27] The human consensus $V_\kappa 4$ carries an exposed tryptophan residue at position

L58 in addition to the conserved Trp L43, which is not quenched in the native state. The denaturation curve is fully reversible, but shows a steep pre-transition baseline followed by a non-cooperative transition. Because of this uncertainy, we report no ΔG_{N-U} value for $V_\kappa 4$, only the midpoint of transition, which is at 1.5 M GdnHCl. For the $V_\kappa 4$ domain Len, a stability of 32 kJ mol^{-1} has been reported.[29]

Analysis of primary sequence and model structures

Large differences are seen in the group of isolated V_H fragments: $V_H 3$ shows the highest yield of soluble protein and thermodynamic stability; $V_H 1a$, $V_H 1b$ and $V_H 5$ show intermediate yield and intermediate or low stability; while $V_H 2$, $V_H 4$ and $V_H 6$ show more aggregation-prone behavior and low cooperativity during denaturant-induced unfolding. The properties of V_κ and V_λ domains are more homogeneous. The thermodynamic stabilities differ by only approximately 10 kJ mol^{-1} in the group of V_κ and in the group V_λ domains. In general, the stability and soluble yield is higher in isolated V_κ domains than in V_λ domains. To analyze possible structural reasons for this different behavior of the variable antibody domains, we analyzed the primary sequence and the modeled structures of the seven human consensus V_H and V_L domains. These models have been published[17] (PDB entries: 1DHA ($V_H 1a$), 1DHO ($V_H 1b$), 1DHQ ($V_H 2$), 1DHU ($V_H 3$), 1DHV ($V_H 4$), 1DHW ($V_H 5$), and 1DHZ ($V_H 6$)) and V_L domains (PDB entries: 1DGX ($V_\kappa 1$), 1DH4 ($V_\kappa 2$), 1DH5 ($V_\kappa 3$), 1DH6 ($V_\kappa 4$), 1DH7 ($V_\lambda 1$), 1DH8 ($V_\lambda 2$), 1DH9 ($V_\lambda 3$)). The quality of the models varies for the different domains. Many antibody structures in the Protein Data Bank (PDB) use, for example, the $V_H 3$ framework, and the chosen template structure for building the model shares 86% sequence identity, excluding the CDR-H3 region (PDB entry: 1IGM) and the structural differences between templates could be traced to distinct sequence differences. In the case of $V_H 6$, the closest templates were human $V_H 4$ and murine $V_H 8$ domains, since no crystal structure of a member of the $V_H 6$ germline family is available in the PDB. Both germline families encode a different framework 1 structural subtype (I) than $V_H 6$ (III).[30] The chosen template for $V_H 6$ (PDB entry: 7FAB) shares 62% sequence identity, excluding the CDR-H3 region and belongs to human $V_H 4$. For each domain model, experimental structures of the ten most closely related domains (human, murine or engineered) available at the time were superimposed by a least-squares fit of the C^α coordinates of residues L3–L6, L20–L24, L41–L47, L51–L57, L78–L82, L89–L93, L102–L108 and L138–L144 in the V_L domain or H3–H6, H20–H24, H41–H47, H51–H57, H78–H82, H89–H93, H102–H108 and H138–H144 in the V_H domain, representing the structurally least variable positions. The aligned structures were scrutinized for the potential structural effects of sequence differences, if necessary consulting additional structures, before templates were assigned for modeling. Currently, experimental structures of V_H domains belonging to human germline families $V_H 1$, $V_H 3$ and $V_H 4$ can be found in the PDB. Modeling of the V_L domains was easier, since for all of the four V_κ and three V_λ families representative structures can be found, although many of them only in the form of light chain dimers (Bence–Jones proteins), not as scFv, Fv or Fab fragments. The analyses of structural subtypes, core packing pattern, conserved hydrogen bonding pattern, charge interactions, influence of mutations on stability etc. are on the basis of the analysis of all antibody variable domains with resolutions better than 3 Å represented in the PDB at the time of the analysis†. We wanted to address three questions regarding the domains in isolation: why is $V_H 3$ so extraordinarily stable; why do $V_H 2$, $V_H 4$ and $V_H 6$ behave comparatively poorly concerning expression and aggregation; and why did V_κ domains give higher yields and are more stable than V_λ domains?

Salt-bridges

Salt-bridges between oppositely charged amino acid residues and repulsions between equally charged amino acid residues play an important role in protein stability.[31] Figure 4(a) shows a representation of an scFv fragment consisting of a $V_\kappa 3$ and $V_H 3$ domain with its characteristic secondary structure. In Figure 4(b), positively charged residues at pH 7.0 are colored in blue and negatively charged residues are colored in red. There is an accumulation of charged residues at the base of each domain. A hydrogen bond from the side-chain OH of tyrosine H104 to the main-chain carbonyl group of H100 orients residue H100, which forms a highly conserved buried salt-bridge with residue H77 (Figure 5(a)). In $V_H 3$, the two terminal nitrogen atoms of the Arg side-chain form a hydrogen bond to the two oxygen atoms of the side-chain carboxylate group of Asp H100. Arg H45 interacts with H100, as well as with Glu H53 and with the side-chain of Tyr H100. The more accessible and less highly conserved residues H99 and H97 complete the charge cluster in a germline family and subtype-dependent manner, with Glu H99 either interacting with an Arg in H97, as seen in the structure with PDB entry 1IGM, or with Arg H45 as seen in structures with the PDB entries 1BJ1, 1INE, 2FB4 and 1VGE. Gln H77 in $V_H 5$ can form only a single hydrogen bond to Asp H100. In addition, to compensate for the shorter Gln side-chain, the side-chain Asp H100 has to move slightly towards H77 and can no longer accept a hydrogen bond from Arg H45,

---

† Presented in the AAAAA website at http://www.biochem.unizh.ch/antibody
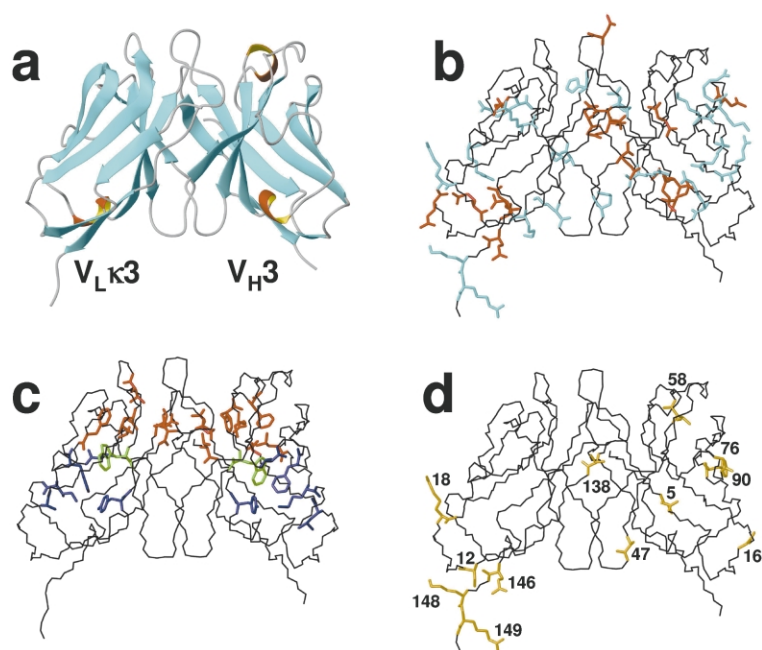
Figure 4. Model structure of an scFv fragment consisting of human consensus $V_κ3$ (PDB entry: 1DH5[17]) and V_H3 domain (PDB entry: 1DHU[17]). (a) Secondary structure with $V_κ3$ on the left and V_H3 on the right side (b) Colored for charged residues (blue: Arg, Lys and His; red: Asp and Glu). At the base of each domain is an accumulation of charged residues, the charge clusters of V_L and V_H domains. (c) Hydrophobic core residues: above the conserved Trp43 (green) is the upper core (red) and below the lower core (blue), see the text for details. (d) Positions possibly influencing folding efficiency are shown in orange; see the text for details. All images were generated using the program MOLMOL.[33]
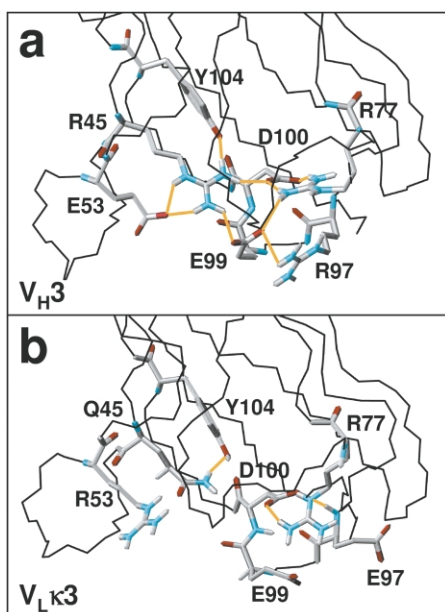
opening up the network of charge interactions and hydrogen bonds.

In V_L domains (Figure 5(b)), the amino acid residue at position L45 is an uncharged Gln and those in positions L53 and L97 are either reversed compared to the amino acid residues at these positions in V_H domains or are uncharged. Instead of two hydrogen bonds connecting the Arg L77 and Asp L100 side-chains, Arg L100 interacts frequently with both Asp L100 and Glu L99, forming a single

Figure 5. Detailed view of the charge cluster of the human consensus (a) V_H3 (PDB entry: 1DHU[17]) and (b) $V_κ3$ (PDB entry: 1DH5[17]) family with hydrogen bonds in orange. Images were generated using the program MOLMOL.[33]

hydrogen bond to each of the two side-chains. In addition, the side-chain of Arg L77 can form a hydrogen bond to the main-chain carbonyl group of Glu L97 (Figure 5(b)). The least stable $V_κ$ domain $V_κ2$ carries Leu at position L45, which is unable to form the side-chain to side-chain hydrogen bond to Tyr L104 conserved in the other V_L domains and in V_H domains (Figure 5(a) and (b)).

Hydrophobic core packing

Another important stabilizing factor is hydrophobic core packing.[32] All model structures were checked for cavities, which would indicate imperfect packing, leading to fewer van der Waals interactions and reduced thermodynamic stability. A van der Waals contact surface was generated for a water radius of 1.4 Å with the program MOLMOL (data not shown).[33] When cavities were found, the surrounding residues were checked whether they would contribute hydrophobic surface area to the cavity. A cavity lined with hydrophobic residues would be less desirable, as a water molecule would be energetically unfavorable at such a position. On the basis of these cavities and sequence comparisons between the different variable domain frameworks, we identified positions in the hydrophobic core that may lead to sub-optimal packing. An overview of the analyzed core residues is given in Figure 4(c). The core residues can be divided into two regions: the upper and lower core according to the orientation shown in Figure 4(a). A layer of invariant residues, including the core Trp in position 43, the conserved disulfide bridge between Cys23 and Cys106, and Gln/Glu6 separate the upper from the lower core. The upper core is built of buried residues above (towards the CDRs) this central core of residues.

Table 2. Sequence alignment of the human consensus V_H and V_L domains at regions possibly influencing thermodynamic stability

AHo[a]	Charge cluster						Upper core										Lower core				
	45	53	77	97	99	100	2	4	25	29	31	41	80	82	89	108	19	74	78	93	104
V_H3	R	E	R	R	E	D	V	L	A	F	F	M	I	R	L	R	L	V	F	M	Y
V_H1a	R	E	R	R	E	D	V	L	A	G	F	I	I	A	A	R	V	F	V	L	Y
V_H1b	R	E	R	R	E	D	V	L	A	Y	F	M	M	R	A	R	L	F	V	L	Y
V_H5	R	E	Q	K	S	D	V	L	G	Y	F	I	I	A	A	R	L	F	V	W	Y
V_H2	R	E	R	D	V	D	V	L	F	F	L	V	I	K	V	R	L	L	L	M	Y
V_H4	R	E	R	T	A	D	V	L	V	G	I	F	I	V	F	R	L	L	V	L	Y
V_H6	R	E	R	T	E	D	V	L	I	D	V	F	I	P	F	R	L	V	I	L	Y
$V_κ1$	Q	K	R	Q	E	D	I	M	A	Q	I	L	G	G	F	Q	V	V	F	I	Y
$V_κ2$	L	Q	R	E	E	D	I	M	S	Q	L	L	G	G	F	Q	A	V	F	I	Y
$V_κ3$	Q	R	R	E	E	D	I	L	A	Q	V	L	G	G	F	Q	A	V	F	I	Y
$V_κ4$	Q	K	R	Q	E	D	I	M	S	Q	V	L	G	G	F	Q	A	V	F	I	Y
$V_λ1$	Q	K	R	Q	E	D	I	L	G	S	I	V	G	K	A	Q	V	V	F	I	Y
$V_λ2$	Q	K	R	Q	E	D	I	L	G	S	V	V	G	K	A	Q	I	V	F	I	Y
$V_λ3$	Q	V	R	Q	E	D	I	L	G	–	L	A	G	N	A	Q	A	I	F	I	Y

[a] Numbering is according to the structurally based scheme described by Honegger & Plückthun.[23]

While most of the CDR side-chains face outwards, a few CDR residues, particularly in CDR-1 and 2, are involved in the upper core, with the consequence that different CDRs have a strong influence on the packing of the upper core (and its contribution to the overall stability), and *vice versa*, the residues of the upper core have an influence on the conformation of the CDRs (and affinity or specificity of antigen binding).[34]

Upper core

The residues 2, 4, 25, 29, 31, 41, 80, 82, 89, and 108 form the upper core. In the sequence alignment shown in Table 2 these residues have been compared for the variable domains. In V_H domains, two sequence motifs can be distinguished: the V_H3-like motif with two bulky aromatic residues at positions H29 and H31 (V_H1b, V_H3, V_H5), the alternative location of the aromatic residues at H25 and H29 (V_H2) and the V_H4/V_H6 motif with Trp at position H41 and a big aliphatic residue at position H25. Figure 6(a) shows a superposition of V_H4 on V_H3, highlighting the differences between these motifs. In the V_H3-like motif Phe H29 and Phe H31 fill the space between the neighboring residues H2, H25 and H108. In the V_H4/V_H6 motif, these two residues are changed to smaller residues. Here, Trp H41 and the methyl group of Val H25 fill up the empty space. V_H1a belongs to the V_H3-like motif but has Gly instead of Phe at position H29. No other residue compensates for this empty space, which results in a hydrophobic cavity (Figure 6(b)). V_H1a, V_H1b and V_H5 have Ala instead of Leu (V_H3) at position H89. There is no obvious compensation for this loss of an isopropyl group. In addition, the substitution of Ala H25 (V_H3) to Gly in V_H5 (Table 2) equals the loss of a methyl group, further weakening the packing of the upper core of V_H5 (Figure 6(c)).

Figure 6(d) shows the superposition of the upper core of the $V_κ3$ and $V_λ1$ domain as representatives of $V_κ$ and $V_λ$ domains. The packing density of the $V_κ$ domains compared to the V_H domains is smaller, because there is only one bulky aromatic
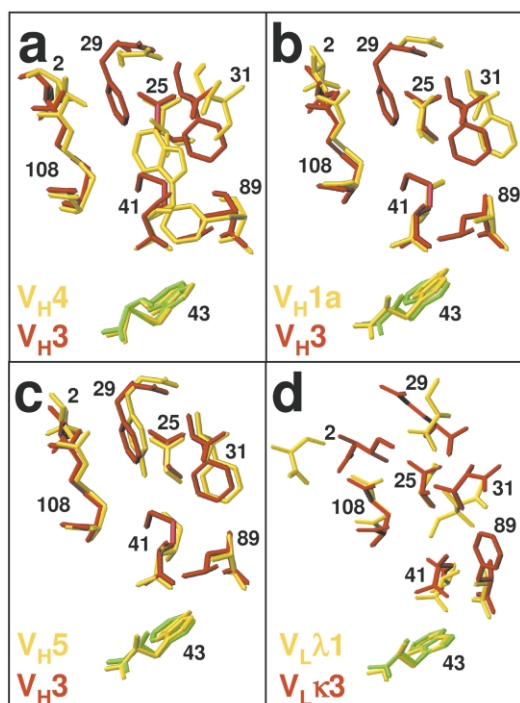
Figure 6. Detailed view of the upper core residues. Superposition of (a) V_H4 (PDB entry 1DHV[17]), (b) V_H1a (1DHA[17]) and (c) V_H5 (1DHW[17]), each in orange, with V_H3 (1DHU[17]) in red and (d) $V_λ1$ (1DH7[17]) in orange with $V_κ3$ (1DH5[17]) in red; see the text for details. The conserved Trp43 is shown in green or yellow. Residues 4, 80 and 82 are not shown, as they do not contribute to the packing differences discussed in the text. Images were generated using the program MOLMOL.[33]

Table 3. Key residues of the human V_H family consensus sequences

AHo[a]	Structural subtype[30]	Residues defining framework class I			Residues differing between well and poorly behaved V_H domains					
		6	7	10	5	16	47	58	76	90
V_H3	II	E	S	G	V	G	A	I	G	Y
V_H1a	III	Q	S	A	V	G	A	I	G	Y
V_H1b	III	Q	S	A	V	G	A	I	G	Y
V_H5	III	Q	S	A	V	G	M	I	G	Y
V_H2	I	E	S	P	K	T	P	I	T	V
V_H4	I	E	S	P	Q	S	P	I	S	S
V_H6	III	Q	S	P	Q	S	S	T	S	S

[a] Numbering according to the structurally based scheme described by Honegger & Plückthun.[23]

amino acid residue in the upper core of V_κ domains at position L89, compared to V_H domains, which have at least two aromatic residues (Table 2). The packing density is further lowered in V_λ domains because of the smaller Gly in position L25 and Ala in position L89 instead of Ala/Ser and Phe, respectively, which are found in V_κ domains (Figure 6(d), Table 2), consistent with a lower thermodynamic stability of V_λ domains.
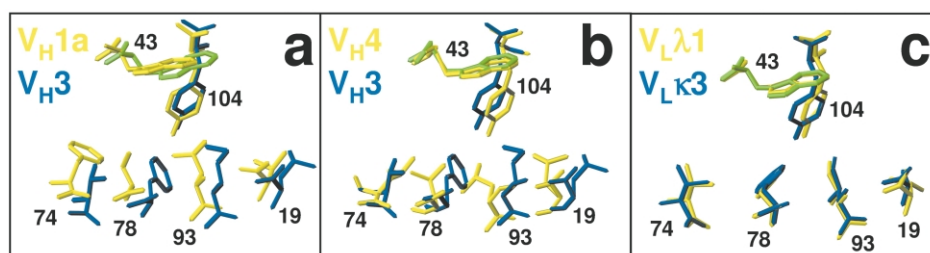
Lower core

Human V_H domains can be classified into three, and murine into four distinct structural subtypes.[30,35] The most discernible predictors of the structural subtype in the sequence of V_H domains are the identities of residues H6, H7 and H10, which determine the conformation of framework 1 directly. Subtype I is characterized[30] by a glutamate residue in position H6, a non-Pro residue in position H7 and a *trans*-Pro in position H10; subtype II by H6 Glu, H7 non-Pro and H10 Gly; subtype III by H6 Gln, H7 non-Pro and any residue in H10; while subtype IV, not represented amongst human germlines, is characterized by H6 Gln, H7 Pro and any residue in position H10 (Table 3). Correlated sequence differences define differences in the lower core packing for the different structural subtypes.[35] Out-of-context mutations within residues H6–H10 frequently introduced by PCR primers used for amplification and, to a lesser degree, mismatches between FR-1 subtype and

lower core packing were found to lead to severe destabilization of the domain.[36]

Within V_H domains, an interesting correlation is seen between stability and framework 1 classification described by Honegger & Plückthun.[30] The most stable human V_H3 domain belongs to structural subtype II, as do several exceptionally stable murine V_H domains and the camelid V_{HH} domains. Human V_H1a, V_H1b and V_H5 with intermediate properties belong to subtype III (Table 3). The V_H domains showing a great tendency for aggregation and non-cooperative denaturation, V_H2, and V_H4, belong to subtype I. V_H6 is a member of subtype III because of its Gln at position H6 and the absence of Pro from position H7. However, previous experiments have shown that the Pro in position H10 might destabilize the domain.[36]

The lower core comprises residues 13, 19, 21, 45, 55, 74, 77, 78, 91, 93, 96, 100, 102, 104, and 145. Residues H19, H93, H78, and H74 (Table 2) occupy adjacent positions in different strands of the outer β-sheet of the domain (Figure 7). They are highly conserved within, but different between different human germline families. Either position H74 or H78 carries an aromatic residue in the favorable frameworks V_H3, V_H5, V_H1a and V_H1b. In contrast, in V_H2, V_H4 and V_H6, both positions are aliphatic. V_H5 has an additional exchange from Met to Trp at position H93. This additional aromatic residue in V_H5 could help compensate for the loss of Phe H78 and the poor interactions in the charge cluster (see above). Apart from Tyr H104, no additional aromatic residue stabilizes the lower

Figure 7. Detailed view of the lower core residues that correspond to framework 1 classification. Superposition of (a) V_H1a (1DHA[17], orange) and V_H3 (1DHU[17], blue), (b) V_H4 (1DHV[17], orange) and V_H3 (1DHU[17], blue) and (c) $V_\lambda1$ (1DH7[17], orange) and $V_\kappa3$ (1DH5,[17] blue); see the text for details. The conserved Trp43 is shown in green. Images were generated using the program MOLMOL.[33]

Table 4. Sequence alignment of the human consensus V_L families

AHo[a]	12	18	138	146	148	149
$V_\kappa 1$	S	R	T	E	K	R
$V_\kappa 2$	P	P	T	E	K	R
$V_\kappa 3$	S	R	T	E	K	R
$V_\kappa 4$	A	R	T	E	K	R
$V_\lambda 1$	S	R	V	T	L	G
$V_\lambda 2$	S	S	V	T	L	G
$V_\lambda 3$	S	T	V	T	L	G

[a] Numbering according to the structurally based scheme described by Honegger & Plückthun.[23]

core of V_H2, V_H4, and V_H6 (Figure 7(b)). FR-1 residue H13, is Val in V_H2, 3, 4 and 6 and Lys with a solvent-exposed ε-amino group in V_H1 and V_H5. The result of these differences is that the entire row of side-chains seems to lean towards Gly H10 and the framework 1 kink in structural subtype II (V_H3), while it is in some intermediate orientation in subtype III (V_H1, V_H5, and V_H6) and points away from the Pro in H10 towards residues 74 in subtype I (V_H2 and V_H4) (Figure 7 (a) and (b)).

V_L domains do not show a similar variability in lower core packing. As a consequence, the lower core residues of V_κ and V_λ domains are almost the same and have similar orientations (Table 2). Their lower core packing most closely resembles that of V_H3 (Figure 7(c)). Despite the sequence differences, all domains seem to achieve quite tight packing of the lower core, without obvious cavities or packing defects and with a very similar core volume and buried hydrophobic surface area.

Residues possibly influencing solubility and folding efficiency

We further examined residues that could correlate with poor expression behavior and a great tendency to aggregate due to kinetic rather than thermodynamic reasons.[37] We started our analysis from a sequence alignment of the human consensus V_H domains grouped by V_H with good biophysical properties (V_H1a, V_H1b, V_H3, V_H5) and more aggregation-prone V_H domains (V_H2, V_H4, V_H6) (Table 3).

It was shown previously that mutations of exposed hydrophobic residues do not change the solubility of the native scFv fragment, as determined by salting-out, but have a profound effect on the *in vivo* folding yield.[12] Position H5 is exposed to solvent and therefore the hydrophilic residue Gln or Lys of V_H2, V_H4, and V_H6 might be thought to decrease the aggregation tendency in contrast to the hydrophobic Val in V_H1a, V_H1b, V_H3, and V_H5. Nevertheless, Val was selected out of Val, Gln, Leu, and Glu in the *in vitro* evolution of loop grafted scFv 4D5Flu towards increased stability,[38] possibly indicating the importance of local secondary structure propensity.

V_H2, V_H4 and V_H6 have a non-glycine residue with a conserved positive φ angle at position H16 (Figure 4(d)), which causes an unfavorable local conformation. Structures that have been determined with a non-Gly residue at position H16 (e.g. PDB entries 1C08, 1DQJ, 1F58) indeed show that the positive φ angle is maintained locally, apparently enforced by the surroundings. In contrast, the odd-numbered V_H all have Gly at this position.

For the antibody McPC603 (PDB entries 1MCP, 2MCP), it has been shown by Knappik *et al*. that the exchange of Pro H47 to Ala, adjacent to another Pro at position H48, does not result in better thermodynamic stability, but enhances folding efficiency.[9] V_H2 and V_H4 carry Pro at position H47. In V_H6, the highly conserved hydrophobic core residue Ile is exchanged to Thr at position H58, which buries an unsatisfied hydrogen bond donor.

The odd-numbered V_H domains have Gly at position H76, in contrast to the even-numbered V_H domains, which carry Thr or Ser. In half of the antibody structures determined that are found in the PDB, the residue at this position has a positive φ angle, indicating that glycine could be better at this position.

The semi-buried position H90 of V_H1a, V_H1b, V_H3, and V_H5 is occupied with Tyr, whereas V_H2, V_H4, and V_H6 have Val or Ser. Replacement of exposed hydrophobic groups by more hydrophilic ones can have a marked effect on folding efficiency and thus on production yield without significantly affecting thermodynamic stability.[12] However, since these effects are presumably due to the destabilization of aggregation-prone intermediates or to interference with the nucleation of aggregates rather than to interactions observable in the native structure, the effects of such mutations are extremely hard to predict.

As the V_L domains can be grouped primarily into κ and λ domains, we concentrated our analysis on a comparison between these two groups. At the solvent-exposed C-terminal end at positions L146, L148 and L149, V_κ domains have charged amino acid residues, in contrast to V_λ domains, which have Thr, Leu and Gly, respectively, at these positions (Table 4, Figure 4(d)). In addition, the hydrophilic Thr in position L138 of κ domains is exchanged to the hydrophobic Val in λ domains (Table 4, Figure 4(d)). These exchanges of less hydrophilic residues in V_λ domains might lower the folding efficiency of these domains and may be a contributing factor to the smaller soluble yield compared to V_κ domains.

Proline is an α-helix and β-strand breaker and thus destabilizes those secondary structures. Positions L12 and L18 in V_L domains are both part of a β-sheet structure. Only $V_\kappa 2$ has Pro at both positions, while Ser and Arg, respectively, are the dominant residues at these positions in the other V_L domains (Table 4, Figure 4(d)). Murine Fab 2H1P contains the only V_κ in the PDB with Pro in both positions. Located in the edge strands of

Table 5. Summary of biophysical characterization of scFv fragments

					Midpoint [GdnHCl] (M)	
scFv	CDR-3	Soluble yield[a]	Insoluble content (%)	Oligomeric state[b]	V_H[c]	V_L[c]
H1aκ3	Short/κ-like[d]	11.1 (1.7)	10	m, D, M	1.8	2.8
H1bκ3	Short/κ-like	12.4 (1.9)	20	M	2.4	3.0
H2κ3	Short/κ-like	2.6 (0.6)	90	M	1.5	2.8
H3κ3	Short/κ-like	6.5 (= 1)	30 ± 10	M	2.8[e]	
H4κ3	Short/κ-like	2.6 (0.4)	90	M	2.3	3.0
H5κ3	Short/κ-like	6.5 (1.0)	50	M	2.2	3.0
H6κ3	Short/κ-like	5.2 (0.8)	80	M	1.2	2.6
H3κ1	Short/κ-like	2.6 (0.4)	50	M	2.8[e]	
H3κ2	Short/κ-like	2.6 (0.4)	20	M	2.9	1.6
H3κ3	Short/κ-like	6.5 (= 1)	30 ± 10	M	2.8[e]	
H3κ4	Short/κ-like	5.2 (0.8)	40	M	2.8	2.0
H3λ1	Short/λ-like[f]	7.8 (1.2)	40	D, M	3.0[e]	
H3λ2	Short/λ-like	5.9 (0.9)	10	D, M	2.9[e]	
H3λ3	Short/λ-like	3.9 (0.6)	10	D, M	2.8[e]	

[a] Given in mg l^{-1} of bacterial culture at OD$_{550}$ of 10, and compared to (in parentheses) the soluble yield of H3κ3.
[b] Oligomeric state in 50 mM sodium phosphate (pH 7.0), 500 mM NaCl with: m, multimer; D, dimer; M, monomer.
[c] Within the scFv fragment.
[d] Sequence of CDR-H3 (short, WGGDGFYAMDY)/CDR-L3 (κ-like, QQHYTTPPT).
[e] Only one transition is visible.
[f] Sequence of CDR-H3 (short, WGGDGFYAMDY)/CDR-L3 (λ-like, QSYDSSLSGVV).

their respective β-sheets (Pro L12 in strand A′ of the inner and Pro L18 in stand B of the outer β-sheet), the two Pro do not perturb the structure noticeably.

Expression and protein purification of scFv fragments

After biophysical characterization of isolated human consensus V_H and V_L domains we tested systematic combinations of V_H and V_L in an attempt to understand their mutual influence on biophysical properties and chose the scFv format, in which the V_H domain is linked _via_ a flexible peptide linker to the V_L domain. To limit the number of possible V_H–V_L combinations of 49, we analyzed the scFv fragments with the most stable V_H domain V_H3 combined with each of the seven human consensus V_L domains and, conversely, the most stable V_L domain $V_κ3$ with each of the seven human consensus V_H domains. We wanted to examine if insufficient stability of one domain can be compensated by a stable partner domain and whether the stabilities of the individual domains are additive or even synergistic in the context of a scFv.

All V_H domains within the scFv fragment carry the same CDR-H3, which is derived from the V_H domain of the well expressing antibody 4D5.[17,24] The $V_κ$ and $V_λ$ domains in the scFv fragments carry the κ-like and λ-like CDR-L3, respectively. All scFv fragments could be expressed in soluble form in the periplasm and purified with IMAC, followed by an anion-exchange column. The fragments all showed the expected molecular masses, and purity of the fragments was over 98%, as confirmed by SDS-PAGE analysis (data not shown).

The subsequent measurements were all carried out with freshly purified proteins. To compare the expression yield of the scFv fragments with the different V_H or V_L domains, we additionally isolated the scFvs with a batch method (for details, see Materials and Methods). To test the error inherent in the yield determination the scFv H3κ3 was purified four times independently. The yield of purified H3κ3 was 6.5($\pm$0.2) mg from a one liter bacterial culture normalized to an OD$_{550}$ of 10, which is approximately the final cell density in a shake flask under these conditions. Yields of all scFv fragments tested were normalized to the yield of H3κ3, and were in the range of 2.6– 12.4 mg l^{-1} (Table 5). H1aκ3 and H1bκ3 with 11.1 mg l^{-1} and 12.4 mg l^{-1}, respectively, (1.7 and 1.9-fold the amount of H3κ3), show the highest yield and H2κ3, H4κ3 and H6κ3 show the lowest yield of scFv fragments with the $V_κ3$ domain with 0.6, 0.4 and 0.6-fold that of H3κ3, respectively. All scFv fragments with V_H3 but different V_L domains show yields only below that of H3κ3, with the exception of H3λ1. The percentage of insoluble protein was determined for H3κ3 in four independent measurements to be 30($\pm$10)%. The other scFv fragments tested show a percentage of insoluble protein between 50% and 10%, with the exception of H2κ3, H4κ3 and H6κ3, which show a percentage of insoluble protein between 80% and 90% (Table 5).

Analytical gel-filtration of scFv fragments

H3κ3 elutes from an analytical gel-filtration column Superdex-75 at a protein concentration of 5 μM in 50 mM sodium phosphate (pH 7.0), 500 mM NaCl with an apparent molecular mass of
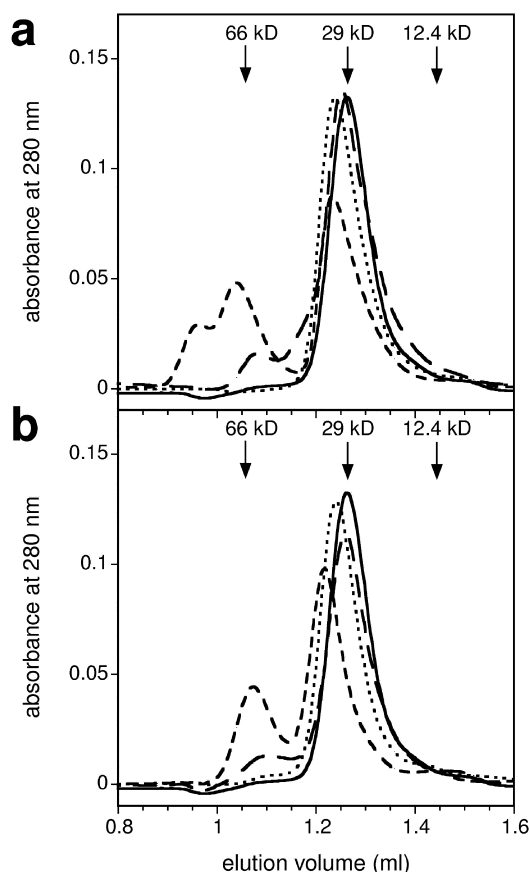
Figure 8. Analytical gel-filtration of scFv fragments (5 μM) on a Superdex-75 column in 50 mM sodium-phosphate (pH 7.0), 500 mM NaCl: (a) H3κ3 (continuous line), H4κ3 (long-dash line), H1aκ3 (short-dash line) and H1aκ3 in the presence of 1 M GdnHCl (dotted line). (b) H3κ3 (continuous line), H3κ1 (long-dash line), H3λ1 (short-dash line) and H3λ1 in the presence of 1 M GdnHCl (dotted line). Arrows indicate elution volumes of molecular mass standards: bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome *c* (14 kDa).

29 kDa, which indicates that H3κ3 is monomeric in solution. The other scFv fragments with $V_κ3$ as the V_L domain are also monomeric under these conditions, with the exception of H1aκ3, which shows minor amounts of dimer and multimer. H4κ3 gives rise to a small amount of dimer of less than 10%. Figure 8(a) shows the chromatogram of H3κ3 as an example for monomeric scFv fragments, along with H1aκ3 and H4κ3. The scFv fragments with V_H3 and a $V_κ$ domain are all monomeric, whereas H3κ1 shows in addition a small dimer peak (Figure 8(b) with H3κ3 as an example for monomeric scFv fragments and H3κ1). In contrast, the scFv fragments with $V_λ$ domains all show monomer–dimer equilibria, with a dimer content from 20% in the case of H3λ1 to 70% in the case of H3λ2 (Figure 8(b), with H3λ1 as an example for scFv fragments with a $V_λ$ domain). With 1 M GdnHCl in the elution buffer, all the scFv fragments that had a dimer fraction under native conditions elute in a single peak at an apparent mass

of 29 kDa, indicating that they are now fully monomeric. The chromatogram in 1 M GdnHCl is shown in Figure 8(a) for H1aκ3 and in Figure 8(b) for H3λ1 as an example for scFv fragments with $V_λ$ domain. It should be noted that this concentration is below the major transition of all scFv fragments. The only exception was H3λ2, which still has dimer content of 20% in 1 M GdnHCl. With 2 M GdnHCl, H3λ2 shows only a monomer peak (data not shown).

Equilibrium unfolding experiments of scFv fragments

Unfolding and refolding of the scFv fragments as a function of denaturant concentration under non-reducing conditions was monitored by the shift of the maximum of the fluorescence emission after excitation at 280 nm. Each scFv fragment shows reversible unfolding behavior (data not shown). The denaturation of the scFv fragments is usually not a two-state equilibrium,[13] because the scFv fragments are built from two domains, which may have different intrinsic stabilities and interact over an interface region and potentially can stabilize each other. Therefore, no ΔG_{N-U} value is reported; instead, the midpoints of the transitions of denaturation are given, which are a semi-quantitative measure for the stability of the scFv fragments. The assignment of the transitions to the V_H or V_L domain results from the determination of the transition of single domains (Table 1). In Table 5, the midpoints are listed for the V_H and V_L domain within the scFv fragments. If only one transition is visible, the midpoint is assigned to both the V_H and V_L domain.

With the knowledge of the denaturation properties of the isolated V_H and V_L domains, and the combinations of these domains in the scFv fragments, it is now possible to study systematically the influence of the interface interaction on the stability of the scFv fragments. Different cases can be distinguished:[14] if the stability of the isolated V_H and V_L domains is very similar, the resulting scFv has the same stability (see Figure 9(a), with H5κ3 as an example). If one domain is significantly more stable than the other, the less stable one can be stabilized through the interface interaction with the other domain (see Figure 9(b), with H1aκ3 with the more stable $V_κ3$ stabilizing V_H1a; and Figure 9(c), with H3κ1 with the more stable V_H3 stabilizing $V_κ1$). Nevertheless, it is possible that, although the stability of the domains is different, almost no stabilization of the less stable domain occurs (see Figure 9(d), with H3κ2 as an example).

The scFv fragments with $V_λ$ domains show an interesting behavior (Figure 10(a), with H3λ1 as an example) because, on the basis of the unfolding transition followed by the shift in the center of spectral mass, the scFv fragments seem to be even more stable than any of the single isolated domains. Apparently, the interface interaction between V_H and V_L is so strong that the domains
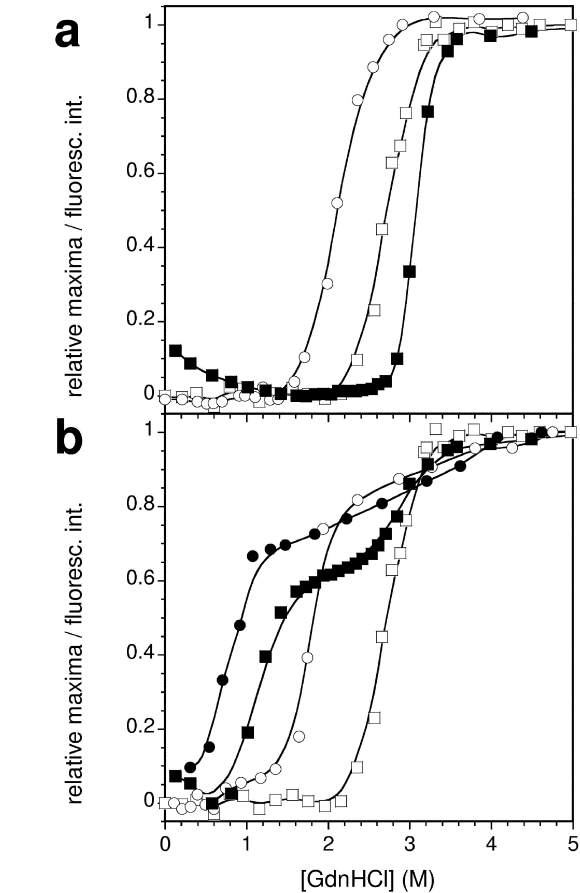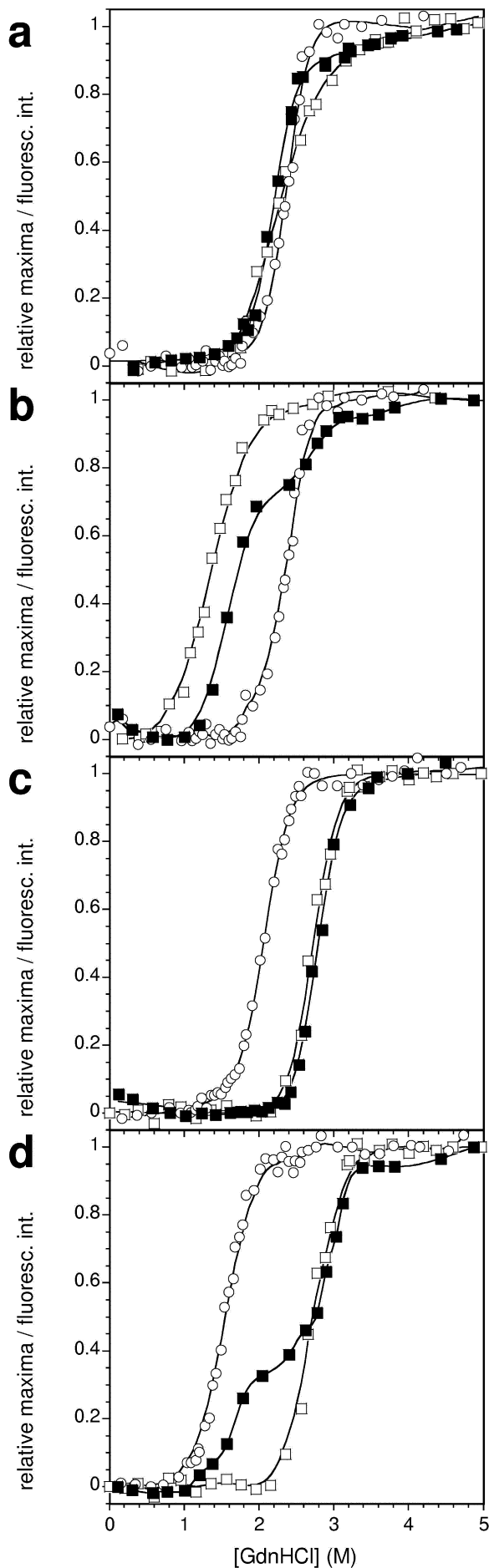
Figure 10. Overlay of GdnHCl denaturation curves to illustrate the role of different CDR-L3 in interface stabilization in V_λ domains. In (a) with H3λ1 with the λ-like CDR-L3 and (b) with H3λ1 with the κ-like CDR-L3 the scFv fragments (filled squares) and constituent isolated V_H3 (open squares) and $V_\lambda1$ (open circles) domains are shown. As the isolated V_λ domains with the κ-like CDR-L3 show non-reversible behavior, in (b) the renaturation curve of $V_\lambda1$ is also shown (filled circles). All unfolding transitions were measured by following the change in emission maximum (in the case of scFv fragments and V_H domains) or fluorescence intensity (in the case of V_L domains) as a function of denaturant concentration at an excitation wavelength of 280 nm. Relative maxima or relative intensities refers to a scaling in which the lowest value is set to 0, the highest to 1. Note that this procedure does not flatten the pre-transition or post-transition baseline.

Figure 9. Overlay of GdnHCl denaturation curves to illustrate different cases of interface stabilization. In each panel the scFv fragment (filled squares) and accompanying isolated V_H (open squares) and V_L (open circles) domains are shown. All unfolding transitions in (a) with H5κ3, (b) with H1aκ3, (c) with H3κ1 and (d) with H3κ2 were measured by following the change in emission maximum (in case of scFv fragments and V_H domains) or fluorescence intensity (in case of V_L domains) as a function of denaturant concentration at an excitation wavelength of 280 nm. Relative maxima or relative intensities refers to a scaling in which the lowest value is set to 0, the highest to 1. Note that this procedure does not flatten the pre-transition or post-transition baseline.

are stabilized above the intrinsic stability of the isolated domains. If the interface finally breaks up, the now isolated domains in the scFv unfold directly, explaining the steep transition. This extraordinary behavior depends strongly on the sequence of CDR-L3.

We constructed and purified the V_λ domains with the κ-like CDR-L3. The isolated V_λ domains with the κ-like CDR-L3 gave very poor yields. They do not show reversible behavior in denaturant-induced equilibrium denaturation and have lower midpoints of denaturation than the corresponding V_λ domain with the λ-like CDR-L3. The combinations of V_H3 with V_λ domains carrying the κ-like CDR-L3 to scFv fragments show yield and dimer to monomer ratios in analytical gel-filtration similar to those for scFv fragments carrying the λ-like CDR-L3 (data not shown) but a different behavior in GdnHCl denaturation. As an example, Figure 10(b) shows H3λ1 with a κ-like CDR-L3, where the $V_\lambda1$ domain is stabilized only slightly in comparison to the renaturation curve of the isolated $V_\lambda1$, indicating that the interface stabilization in this case is not so strong. It should be noted that the only difference between the two scFv fragments in Figure 10(a) and (b) is the different CDR-L3, which obviously causes this dramatic stabilization difference. The κ-like CDR-L3 with proline in position 136 builds a rigid Ω-loop, which probably interferes with the perfect orientation between V_H and V_L.

In summary, the most stable scFv fragments found to denature only starting above 2 M GdnHCl are H3κ3, H1bκ3, H5κ3 and H3κ1. Although the isolated V_λ domains are rather unstable by themselves, in combination with V_H3 they can build very stable scFv fragments, but depend on the CDR-L3 for this effect. Most likely this CDR is responsible for a favorable orientation of V_L to V_H and thus enables a tighter interaction through the interface. ScFv fragments with an intermediate stability starting denaturation above 1 M GdnHCl are H1aκ3, H2κ3, H3κ2 and H3κ4, while H4κ3 and H6κ3 are scFv fragments with a modest stability, starting denaturation under 1 M GdnHCl.

Discussion

Antibodies are prototypical multi-domain proteins. IgGs, Fabs and even scFv fragments thus frequently do not show fully cooperative unfolding behavior of all domains at once, but show complex multi-stage unfolding curves as a function of denaturant concentration in equilibrium unfolding experiments. The more domains that are present in a construct, the more difficult it becomes to interpret these curves in terms of functional stability. The stability of any of the fragments, be they scFv, Fab or full IgG, is usually limited by the least stable of the domains and by the strength of its interaction with the surrounding domains. The weakest of the V_H and V_L domains present are

clearly limiting in scFv fragments, and in Fabs, and presumably in full IgGs. On the other hand, very stable V domains may no longer be limiting, as they can become more stable than the weakest of the constant domains. To improve any suboptimal antibody construct by rational engineering, the "weakest link" has to be identified and improved. To assess the contribution of individual domains to the overall stability of a fragment reliably, we have compared the stability of the isolated V_L and V_H domains to that of the scFv made up of these domains. The scFv is the simplest system to study the contribution of the domain interactions to stability. The domains are forced to be present in stoichiometric ratios and their interaction is concentration-independent. While the contributions of the CDR regions to the intrinsic stability of the domains and to their mutual stabilization though the V_L/V_H dimer interface make a prediction of the stability of an antibody fragment difficult, knowing the contribution of the different consensus frameworks to the overall stability facilitates the identification of the most probable causes of insufficient stability in a given framework greatly, and allows us to devise generally applicable solutions for the improvement of the less stable frameworks.

Classification of scFv fragments

In this study, we analyzed for the first time in a systematic way the biophysical properties of the human V_H and V_L domains in the form of germline family consensus sequences and the combinations of the most stable V_H domain V_H3 with all V_L domains and the most stable V_L domain $V_\kappa3$ in combination with all V_H domains in the scFv format. As the fragments represent the major families of human variable antibody domains, each in the form of one consensus sequence, we have now a reference system for assessing the properties of all human antibodies. Moreover, by this systematic combination of domains in scFv fragments, we can now understand the intrinsic properties of the domains in this assembly as well as their contributions due to domain interactions. The scFv fragments examined can be classified by their properties. There is a very strong correlation between the properties of the isolated domains and their combination in a scFv fragment.

H3κ3 is the most favorable scFv fragment with superior properties: the scFv fragment as well as the isolated V_H3 and $V_\kappa3$ domains show a high thermodynamic stability and good expression yield. The scFv fragment is monomeric in solution and behaves well under all circumstances.

H1bκ3, H5κ3 and H3κ1 have only good properties with high stability, even better expression yield and monomeric state in solution, even though the variable domains themselves are not as favorable as V_H3 and $V_\kappa3$.

H1aκ3 shows a good yield and an intermediate stability, as expected from the isolated V_H

fragment. Under native conditions, this scFv with the CDR-3s used is not a monomer, which is again the same behavior as for the isolated V_H1a fragment.

H2κ3, H4κ3 and H6κ3 give comparatively low yields of soluble proteins in periplasmic expression. They are monomeric in solution but under the standard conditions used give rise to 80% insoluble scFv in the periplasm. This is obviously caused by the V_H domains, and the isolated V_H domains were totally insoluble when expressed by themselves. These scFv fragments show low or intermediate stability, as expected from the isolated V_H domains. However, we have shown recently that they can be engineered to biophysical properties equal to H3κ3.[39]

H3κ2 and H3κ4 show intermediate yield and stability, and are monomeric in solution.

H3λ1, H3λ2 and H3λ3 show good to intermediate yield but display dimer formation under native buffer conditions. Although the stability of the isolated V_λ domains is low, with V_H3 these λ domains build up the most stable scFv fragments tested here. This shows that there is a particularly strong interaction between V_L and V_H in case of V_λ domains, which more than compensates for the intrinsic low stability of V_λ domains. This stability is strongly dependent on the sequence of CDR-L3, because the κ-like CDR-L3 with the *cis*-proline residue at position 136 in the scFv fragments lead to only a small stabilization of the V_L domain through the interface interaction with V_H3.

Prediction of scFv properties and role of CDR-3

One practical consequence of the present study is the ability to make predictions of properties of scFv fragments on the basis of the germline families of V_H and V_L domains. Even though we can now rationalize many properties and extrapolate to the behavior of combinations not mentioned in this study, we noticed the rather important influence of the CDR-3 sequence. The CDR-3 of both variable domains is a significant part of the interface between V_H and V_L, and thus influences the orientation between these domains. In this study, we want to focus on the influence of the family-specific changes in the framework. Therefore, we chose the same CDR-3 within V_H, V_κ and V_λ domains. To fully describe the system of antibodies, a detailed analysis of the role of CDR-3 sequence and length will have to be done next.

In the isolated variable domains, the CDR-3 also has an important influence. In the case of the V_H domains the 17 amino acid residue long CDR-H3 renders the V_H domains soluble, enabling the analysis of these domains. The V_H domain of the murine anti-c-*myc* antibody 9E10,[40] which has an extended CDR-H3, and the V_H domains of heavy chain antibodies (V_{HH}) naturally occurring in Camelidae[41] show an increased solubility. X-ray structures of V_{HH} domains show that the extended

CDR-H3 covers partially the region equivalent to the interface to V_L,[42,43] which implies that the long CDR-H3 we found could cover the hydrophobic interface in a similar way. The V_H3 domain was the only V_H domain that could be purified with the short CDR-H3, further demonstrating the superior properties of the V_H3 framework. The free energy of unfolding of the V_H3 domain with the shorter CDR-H3 was 13 kJ mol^{-1} lower than that with the long CDR-H3. This may be explained by additional interactions of the longer CDR-H3 with the residues of the hydrophobic interface. The isolated V_λ domains with the κ-like CDR-L3 show, in contrast to those with the λ-like CDR-L3, no reversible behavior during denaturant-induced denaturation, and display a lower stability. These properties are probably caused by the unusual *cis*-proline residue at position 136 conserved in V_κ domains but not in V_λ domains, causing an Ω-loop conformation of the CDR-L3, which may clash with CDR-L1 in the V_λ framework.

Analyzing the structure

Our analysis has shown that the V_H3 consensus domain alone and in combination with a V_L domain is thermodynamically much more stable than any of the other V_H frameworks, followed by V_H1b and V_H5. Experimentally, the isolated odd-numbered V_H domains can be unfolded reversibly, while the even-numbered V_H domains cannot. Thus, large differences exist in human V_H domains regarding expression and folding, as well as in thermodynamic stability. We then attempted to correlate these experimental data with structural factors. While no single decisive defect was found, an unexpectedly large number of smaller imperfections always seemed to disfavor the even-numbered V_H domains. The superior quality of V_H3 is thus due to a number of small contributions. The differences are seen in all parts of the molecule, in the hydrophobic core and residues involved in ionic interactions, and residues presumably involved in the kinetic branching to folding or aggregation, such as exposed hydrophobic residues, non-glycine residues in positions that require a positive φ angle, and two proline residues previously identified as leading to aggregation-prone behavior. While our analysis is on the basis of structural criteria only at this point, the redundancy of the correlations suggests that they may, at least in part, explain the biophysical behavior. A more detailed quantitative analysis of energetic contributions must await the determination of experimental structures of the domains and the constructions of point mutants. Such a systematic study has now been completed[39] for V_H6, and indeed the analysis presented here could be experimentally verified, and V_H6 was improved to the same level as V_H3.

The properties within the V_L domains are more homogeneous, and the biggest differences are visible between κ and λ domains. V_κ domains of

Table 6. Framework usage *in vivo* and *in vitro*

	Human family	Germline segments[a]	Framework usage (%)		
			137 binders from Griffiths library[b]	Theoretical distribution of HuCAL®[c]	250 binders from HuCAL®[d]
V_H	1a and 1b	24[e]	13	12	16
	2	6	0	9	22
	3	43	74	10	36
	4	22	11	19	1
	5	4	1	18	13
	6	2[f]	0	32	12
V_L	κ1	25	7	16	13
	κ2	12	47	16	5
	κ3	9	2	16	17
	κ4	1[f]	0	16	12
	λ1	9	28	12	13
	λ2	8	4	12	11
	λ3	14	9	12	28
	Other	26	2	–	–

[a] Taken from VBASE (http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html); 51 human germline segments for V_H and 76 for V_L.
[b] Taken from Griffiths *et al.*[44] Originally 215 binders were sequenced but there are only 137 unique sequences. The Griffiths library is built from an *in vitro* rearranged germline bank, therefore the theoretical distribution is given by the percentage of germline segment, present in the human genome, as given in column 3.
[c] Theoretical distribution is corrected for size of sublibraries and percentage of correct clones in the original HuCAL® scFv library.[17]
[d] Taken from Knappik *et al.*[17]
[e] Including DP-21 (V_H7).
[f] One germline segment.

type $V_κ$3 appear to have slight advantages with respect to yield and stability, while within $V_λ$ domains this is true for $V_λ$1 with respect to stability and for $V_λ$2 with respect to yield. In general, the yield of soluble protein and thermodynamic stability is higher in $V_κ$ domains, compared to $V_λ$ domains. This is possibly correlated to the higher packing density in the upper core, which may explain, in part, the higher thermodynamic stability, while the more hydrophilic C terminus of the $V_κ$ domains may increase the solubility and thus the soluble yield.

Comparison to framework usage

Human antibodies directed against a wide variety of antigens can now be obtained from several antibody libraries.[17,44,45] Antibody fragments that bind the desired target molecule can be enriched from these libraries, but additional factors such as expression behavior, toxicity of the expressed antibody construct to the bacterial host, protease sensitivity, folding efficiency, and stability affect the selection. V_H3 holds a dominant position within the V_H frameworks with regard to thermodynamic stability, while the V_L domains are more homogeneous in their behavior. Is this different behavior reflected in the framework usage *in vivo* and *in vitro*?

We first summarize the frequency of subtypes found in the human germline, then critically discuss the difficulty of establishing their frequency in natural antibodies, and we examined two widely used antibody libraries[17,44] and compared the usage with the theoretical distributions of these libaries (Table 6). Out of 51 functional human V_H germline segments,[46] 22 genes (43%) code for the V_H3 family. V_H2 with three sequences and V_H5 and V_H6 with two and one sequences, respectively, are clearly under-represented (Table 6). Out of 76 human V_L germline segments, 25% code for $V_κ$1, one segment codes for $V_κ$4 and the other frameworks are distributed equally, with frequencies between 8% and 14%.

In a subset of the Kabat Database, which contains only sequences more than 90% complete and includes the CDR-3 and FR-4 sequences, 24% belonged to V_H1, 6% to V_H2, 43% to V_H3, 22% to V_H4, 4% to V_H5 and 2% to V_H6. Of the V_L domains, 25% belonged to Vκ1, 12% to Vκ2, 9% to Vκ3, 1% to Vκ4, 9% to Vλ1, 8% to Vλ2 and 14% to Vλ3. However, this is clearly not an unbiased representation of the *in vivo* usage of the different germline segments. Some of the less stable germline families are actually over-represented due to particular interest in disease-related phenotypes: almost all of the $V_κ$4 sequences listed are associated with light chain deposition disease and light chain amyloidosis, as are many of the $V_λ$ sequences. The majority of the V_H6 sequences in the database derive from a single study, in which sequences derived from that germline family were compared in neonatal and adult blood; the V_H6 family was used specifically for this study because it is derived from a single germline segment. Analysis of the distribution of the four different structural subtypes of V_H in human and murine germline and rearranged sequences showed that even though the distribution of their occurrence varies widely between human and murine

germline genes, their frequencies in rearranged genes are much more similar, with type II and type III being most abundant.[30] It is not clear whether this effect is just a coincidence, or whether it results from some functional specialization of the different framework classes, which need to be present in a balanced mixture to ensure the proper function of the humoral immune system.[30]

We next investigated the relationship between biophysical properties and the frequency of subtypes in various antibody libraries. The Griffiths library is built from heavy and light chain germline segments used *in vivo* and is constructed by PCR with synthetic CDR-3 and FR-4 sequences.[44] If the biophysical properties of the V_H and V_L frameworks were equal, the distribution of V_H and V_L germline segments *in vivo* should be reflected in the isolated binders of the Griffiths library. Out of 137 selected binders with unique sequence, 101 binders (74%) use the V_H3 framework, which is 1.7 times more than expected from the usage in germline segments. No binder containing the V_H2 or V_H6 framework was selected in the original experiments and the use of the V_H4 framework was reduced. The use of $V_\kappa2$ (47%) and $V_\lambda1$ (28%) is increased while the other frameworks are used less compared to the germline segment distribution.

The HuCAL® consists of combinations of seven V_H and seven V_L synthetic consensus frameworks connected *via* a linker region forming 49 master genes.[17] Comparing the framework usage of 250 selected binders from HuCAL®,[17] the V_H3 family is selected in 91 cases (36%), which is more than three times the expected frequency, which was calculated with the knowledge of sublibrary size and number of correct clones (Table 6).[17] Only three out of 250 binders use the V_H4 framework, which is far below the expected 19% usage. The V_H6 family usage in the HuCAL® is decreased in comparison to the theoretical distribution. The V_L framework usage is in the expected range, with the exception of $V_\kappa2$, which is used three times less than expected (5%), and $V_\lambda3$, which is used twice more than expected.

Even taking into account the limited number of antigens used in these published experiments, which will bias the use of a family-specific framework, the V_H3 framework is favored significantly in V_H germline segment usage as well as in the two libraries. In the case of the V_L framework usage, no trend could be observed because in the two libraries different V_L domains are preferred. These observations can be confirmed by two recent studies describing the selection of intrabodies in yeast. In the reducing environment of the cytosol the conserved disulfide bonds cannot form and this absence leads to loss of stability.[47] Therefore, intrabodies have to be intrinsically stable to overcome this loss, give rise to sufficient yield of soluble protein and low aggregation tendency in order to show a biological effect.[48] In these studies,[48–50] several intrabodies were selected and all have only V_H3 frameworks, again underlining the superior quality of the V_H3 framework. Visintin *et al.*[49] obtained intrabodies with mainly the $V_\kappa1$ framework, which we show to be the second most stable V_L fragment, while Tse *et al.*[50] obtained an equal distribution of V_κ and V_λ domains. Using the results described here, any framework can now be engineered[39] to have superior properties and thus form the basis of intrabodies, greatly expanding the proposed "intracellular consensus",[49] which was simply the V_H3 signature. This is important, as framework residues both in the "outer loop" and near the pseudo 2-fold axis relating V_L and V_H can contribute significantly to antigen binding, which is the main reason why nature uses so many frameworks.[23]

Recommendations for library design

The present study makes recommendations for future antibody library design optimizing the percentage of antibodies that are thermodynamically stable and give good expression yields. Up to now, there have been several approaches to building antibody libraries: some mimic the natural distribution of antibodies and are built from germline segments rearranged *in vitro* or *in vivo*.[44,45] These libraries use the natural germline segments and are thus biased towards the V_H3 framework. Nevertheless, they contain a large percentage of other framework combinations with less desirable properties.

Another approach is to construct a single framework library,[51–53] which is usually built on the most frequently used germline segment DP47, which belongs to the V_H3 family, but different V_L germline segments (DPL3, DPK22, DPK9) have been preferred by different groups. In this approach, the resulting antibodies should theoretically be stable and give good yields. However, single frameworks have the great disadvantage that, by relying on only one framework, the antigen-binding diversity is reduced significantly. It is well known that important framework residues are located in the interface, where they contribute to the cavity exploited for hapten binding and by side-chains of bound peptides, and they provide the outer loop residues that contribute greatly to affinity and diversity.[23]

The HuCAL®[17] uses another approach, in which germline family-specific consensus sequences are used that should be the optimum regarding stability and expression yield within each family. The fully synthetic nature of the library allows any composition of frameworks to be adjusted and, especially, the improvement of the less ideal frameworks, to maintain both diversity and stability.

Our suggestion for future antibody library designs is therefore to combine all three approaches with the results of our analysis of the properties of the variable domain families: the consensus sequences of V_H family 1, 3 and 5 and

of V_L family $V_\kappa1$, $V_\kappa3$ and $V_\lambda1$ can be taken, because even the isolated domains are thermodynamically stable and the scFv fragments all give reasonable yields. All six domains should be randomized in CDR-1, CDR-2 and CDR-3, using λ-like CDR-L3s without proline at position 136 with $V_\lambda1$. The resulting antibodies should give yields between 3 mg and 10 mg per liter of bacterial culture of A_{550} of 10, provided the absence of deleterious mutations in the CDRs, they should be thermodynamically stable and denature above GdnHCl concentrations of 2 M, and they should give rise to small amounts of insoluble protein during expression. The selected antibodies should be monomeric in solution, with the exception of antibodies having the $V_\lambda1$ domain, which may have monomer–dimer equilibrium, but using the Fab format should give only monomers. Most importantly, they will still cover the necessary diversity in the most efficient manner. The essentials of this design have been realized in HuCAL® GOLD (MorphoSys AG†). Moreover, an even greater diversity of frameworks, which contribute important interactions beside the CDRs, can be obtained by re-engineered, even-numbered V_H domains. We have now found in a model study that V_H6 can be engineered to approximately the same properties as V_H3 with just a few point mutations.[39]

Conclusions

We present here the first systematic evaluation of human variable antibody domains. While within the isolated V_H and V_L domains significant differences in terms of folding yield and/or stability are visible, the differences are much smaller in the interacting combinations making up the scFv format. The interface between V_H and V_L domains has an important influence on stability and yield. First, it covers the hydrophobic residues that are exposed and probably the cause of the lower yields of the isolated domains in some cases. Second, the stability of a less stable domain is increased through the interface interactions with a more stable variable domain. Therefore, a variety of ways lead to scFv fragments with favorable properties, either by using intrinsically stable variable domains or by increasing stability through favorable interface interaction. CDR-3 sequence and length seem to play a critical role, because CDR-3 regions interact directly across the interface.

Materials and Methods

Construction of expression vectors

Starting point for all expression vectors were the scFv master genes of the HuCAL® library in the orientation

† http://www.morphosys.com

V_H-$(Gly_4Ser)_4$-V_L in the expression vector pBS13,[17] which all carried CDR-H3 and CDR-L3 of the antibody hu4D5-8.[24]

The seven isolated human consensus V_H domains were PCR amplified from the master genes and the CDR-H3 region between the *Bss*HII and *Sty*I restriction sites was then exchanged to code for a CDR-H3 found by metabolic selection (J. Burmester *et al.*, unpublished results): YNHEADMLIRNWLYSDV. The final expression plasmids were derivatives of the vector pAK400,[54] in which the expression cassette of the seven different V_H domains had been introduced between the *Xba*I and *Hin*dIII restriction sites, and where the *skp* cassette[55] had been introduced at the *Not*I restriction site. The expression cassette consists of a *phoA* signal sequence, the short FLAG-tag (DYKD), one of the seven V_H domains and a hexahistidine tag.

The seven isolated human consensus V_L domains were cut out from the master genes with the restriction enzymes *Eco*RV and *Eco*RI, and ligated into a pAK400 derivative with these restriction sites. The CDR-L3 of the V_λ domains between the *Bbs*I and *Msc*I restriction sites was exchanged to QSYDSSLSGVV (L107-L138). This λ-like CDR-L3 is a consensus CDR-L3 from sequences found in the Kabat Database[56] for V_λ domains, while the κ-like CDR-L3 was taken from hu-4D5-8 and encodes the conserved *cis*-proline residue in position L136. The chosen length of the consensus λ-like CDR-L3 is found in 20% of the sequences, representing the highest percentage. The tryptophan residue at position L109, which is the most frequent residue with 54%, was exchanged to tyrosine, which is present in 20% of the sequences, to avoid interference with the native state fluorescence signal of the conserved unique tryptophan residue. The final expression cassette consists of a *pelB* signal sequence, one of the seven V_L domains and a hexahistidine tag.

The scFv fragments were cloned *via* the restriction sites *Xba*I and *Eco*RI into the expression plasmid pMX7. The κ-like CDR-L3 was exchanged in the V_λ domains as reported above. The final expression cassette consists of a *phoA* signal sequence, the short FLAG-tag (DYKD), one of the seven V_H domains, a $(Gly_4Ser)_4$ linker and one of the seven V_L domains the long FLAG-tag (DYKDDDD) and a hexahistidine tag.

Soluble periplasmic expression

All expression experiments (for soluble, insoluble and batch purifications) were carried out with *E. coli* SB536.[57] dYT medium (30 ml) containing 30 μg ml^{-1} of chloramphenicol and 1.0% (w/v) glucose was inoculated with a single bacterial colony and incubated overnight at 25 °C. One liter of dYT medium (30 μg ml^{-1} of chloramphenicol, 50 mM K_2HPO_4) was inoculated with the pre-culture and incubated at 25 °C (5 l flask with baffles, 105 rpm). Expression was induced at an OD_{550} of 1.0 by addition of IPTG to a final concentration of 0.5 mM. Incubation was continued for 18 hours, when the cell density reached an OD_{550} between 8.0 and 11.0. Cells were collected by centrifugation (8000*g*, ten minutes at 4 °C), suspended in 40 ml of 50 mM Tris–HCl (pH 7.5), 500 mM NaCl and disrupted by French press lysis. The crude extract was centrifuged (48,000*g*, 60 minutes at 4 °C), the supernatant passed through a 0.2 μm filter and applied directly to IMAC chromatography (see below).

Preparative two-column purification

The proteins were purified using the two-column, coupled in-line procedure.[8] In this strategy, the eluate of an IMAC column, which exploits the C-terminal His tag, was loaded directly onto an ion-exchange column. Elution from the ion-exchange column was achieved with a 0–800 mM NaCl gradient. The V_H and V_κ domains were purified with a HS cation-exchange column in 10 mM Mes (pH 6.0) and the V_λ domains and the scFv fragments with an HQ anion-exchange column in 10 mM Tris–HCl (pH 8.0). Pooled fractions were dialyzed against 50 mM sodium phosphate (pH 7.0), 100 mM NaCl.

Insoluble periplasmic expression

LB medium (30 ml, containing 30 μg ml^{-1} of chloramphenicol, 1% glucose) was inoculated with a single colony and incubated overnight at 37 °C. One liter of SB medium (20 g l^{-1} of Tryptone, 10 g l^{-1} of yeast extract, 10 g l^{-1} of NaCl, 50 mM K$_2$HPO$_4$) with 10 μg ml^{-1} of chloramphenicol, 0.1% glucose, 0.4 M sucrose added was inoculated with 10 ml of the preculture and incubated at 25 °C. Expression was induced at an A_{550} of 0.8 by addition of IPTG to a final concentration of 0.05 mM. Incubation was continued for about 15 hours at 25 °C. After centrifugation, cells were suspended in 100 mM Tris–HCl (pH 8.0), 2 mM MgCl$_2$ and disrupted by French press lysis. Inclusion bodies were isolated following a standard protocol.[58] The inclusion body pellet from one liter of bacterial culture was solubilized at room temperature in 10 ml of solubilization buffer (0.2 M Tris–HCl (pH 8.0), 6 M guanidine hydrochloride (GdnHCl), 10 mM EDTA, 50 mM DTT). The resulting solution was centrifuged and the supernatant dialyzed against solubilization buffer without DTT at 10 °C. The sample was loaded onto a nitrilotriacetic acid column (NTA, Qiagen), which had been charged with Ni^{2+}, and IMAC under denaturing conditions was performed. The eluate was diluted (1:10, v/v) into refolding buffer (0.5 M Tris–HCl (pH 8.5), 0.4 M arginine, 5 mM EDTA, 20% (v/v) glycerol, 0.5 mM ε-amino-caproic acid, 0.5 mM benzamidinium–HCl) at 16 °C at a final protein concentration of 1 μM. The formation of disulfide bonds was catalyzed either by the presence of reduced glutathione (GSH) and oxidized glutathione (GSSG) in the refolding buffer at [GSH]:[GSSG] of 0.2 mM:1 mM or 5 mM:1 mM, as indicated. The refolding mixture was incubated at 16 °C for 20 hours and dialyzed against 50 mM sodium phosphate (pH 7.0), 100 mM NaCl.

Ni-NTA batch purification

A portion (20 ml) of the supernatant of the French press lysate of *E. coli* expressing the scFv fragments was incubated with 2 ml of a 50% (w/v) Ni-NTA slurry for 30 minutes at room temperature. The suspension was placed into an empty column with a diameter of 1.5 cm and washed extensively with 50 mM sodium-phosphate (pH 7.0, 1 M NaCl. To remove unspecific binding proteins, the column was washed with 30 mM imidazole. The scFv fragments were eluted by adding 250 mM imidazole. The purity of the samples was checked by SDS-PAGE analysis and the concentration was determined by measuring absorbance at 280 nm. Four scFv fragments were purified in parallel, and H3κ3 was always processed in parallel as a control. The yield was normalized to the yield of H3κ3, and to a one liter expression culture with an OD$_{550}$ of 10.

Determination of soluble to insoluble protein ratio

An aliquot of a French press lysis extract of a one liter scFv fragment expression experiment was centrifuged at 4 °C for 30 minutes at 16,000g. The supernatant (soluble fraction) and the precipitate (insoluble fraction), which was resuspended in 50 mM Tris–HCl (pH 7.5), 500 mM NaCl, were analyzed by SDS-PAGE followed by Western blot with the anti-His antibody 3D5 as described.[59] Chemiluminiscence was detected using a Chemi-Imager™ 4400 (Alpha Innotech Corporation) and the density of the bands was determined with the software ChemiImager™ 5500 (Alpha Innotech Corporation). As the method involves many steps, the error is possibly high, and therefore we give the values as a percentage of insoluble material, rounded to tens, with an estimated error of 10%.

Gel-filtration chromatography

Samples of purified proteins were analyzed on a gel-filtration column equilibrated with 50 mM sodium phosphate (pH 7.0), 500 mM NaCl. The isolated V_H domains and the scFv fragments at a concentration of 5 μM were injected onto a Superdex-75 column (Pharmacia) and the isolated V_κ domains at a concentration of 50 μM and 5 μM on a Superose-12 column (Pharmacia) in a volume of 50 μl and a flow-rate of 60 μl min^{-1} on a SMART-system (Pharmacia). The V_λ domains were injected onto a silica-based TSK-Gel® G3000SWXL column (TosoH) on a HPLC system (Hewlett–Packard) in a volume of 50 μl at a concentration of 5 μM and a flow rate of 0.5 ml min^{-1}. Lysozyme (14 kDa), carbonic anhydrase (29 kDa) and bovine serum albumin (66 kDa) were used as molecular mass standards. Elution was followed by detection of the absobance at 280 nm in the case of the SMART-system and at 220 nm in the case of the HPLC system.

Ultracentrifugation

Sedimentation equilibria were determined with an XL-A analytical ultracentrifuge (Beckmann) using an AN-60Ti four-place analytical rotor. The samples were dialyzed against 10 mM sodium phosphate (pH 7.0), 100 mM NaCl overnight and loaded into a standard six channel 12 mm path length cell at a sample A_{280} of 0.4. The fluorocarbon FC43 was added to each cell sector to provide a false bottom. The samples were run for 24 hours at 20 °C at 19,000 rpm (29,100 g). Data were collected at 280 nm at a radial spacing of 0.001 cm and a minimum of ten scans were averaged for each sample. Data were analyzed with software provided by the instrument manufacturer, using models that assumed either the presence of a single species or of a monomer–dimer equilibrium as described.[60] Solvent densities and sample partial volumes were calculated using standard methods.

Equilibrium denaturation experiments

Fluorescence spectra were recorded at 25 °C with a PTI Alpha Scan spectrofluorimeter (Photon Technologies, Inc., Ontario, Canada). Slitwidths of 2 nm were used for excitation and emission. Protein/GdnHCl mixtures

(1.6 ml) containing a final protein concentration of 0.3 μM and denaturant concentrations ranging from 0 M to 5 M GdnHCl were prepared from freshly purified protein and a GdnHCl stock solution (7.2 M, in 50 mM sodium phosphate (pH 7.0), 100 mM NaCl). Each final concentration of GdnHCl was determined from its refractive index. After overnight incubation at 10 °C, the fluorescence emission spectra of the samples were recorded from 320 nm to 370 nm with an excitation wavelength of 280 nm. The fluorescence emission maximum, which was determined by fitting the fluorescence emission spectrum to a Gaussian function (isolated V_H domain and scFv fragments), or the fluorescence intensity at 345 nm (isolated V_L domains) was plotted *versus* the GdnHCl concentration. Protein stabilities for the isolated human consensus V_H and V_L domains were calculated as described.[16] To compare V_H, V_L and scFv denaturation curves in one plot, relative emission maxima and fluorescence intensities were scaled by setting the highest value to 1 and the lowest to 0.

Acknowledgements

References

1. Hudson, P. J. (1998). Recombinant antibody fragments. *Curr. Opin. Biotechnol.* **9**, 395–402.
2. Willuda, J., Honegger, A., Waibel, R., Schubiger, P. A., Stahel, R., Zangemeister-Wittke, U. & Plückthun, A. (1999). High thermal stability is essential for tumor targeting of antibody fragments: engineering of a humanized anti-epithelial glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res.* **59**, 5758–5767.
3. Harris, B. (1999). Exploiting antibody-based technologies to manage environmental pollution. *Trends Biotechnol.* **17**, 290–296.
4. Dooley, H., Grant, S. D., Harris, W. J. & Porter, A. J. (1998). Stabilization of antibody fragments in adverse environments. *Biotechnol. Appl. Biochem.* **28**, 77–83.
5. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M. *et al.* (1988). Single-chain antigen-binding proteins. *Science*, **242**, 423–426.
6. Glockshuber, R., Malia, M., Pfitzinger, I. & Plückthun, A. (1990). A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry*, **29**, 1362–1367.
7. Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotny, J., Margolies, M. N. *et al.* (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **85**, 5879–5883.
8. Plückthun, A., Krebber, A., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K. & Riesenberg, D. (1996). Producing antibodies in *E. coli*: from PCR to fermentation. In *Antibody Engineering, A Practical Approach* (Mc Cafferty, J., Hoogenboom, H. R. & Chiswell, D. J., eds), pp. 203–252, Oxford University Press, New York.
9. Knappik, A. & Plückthun, A. (1995). Engineered turns of a recombinant antibody improve its *in vivo* folding. *Protein Eng.* **8**, 81–89.
10. Forsberg, G., Forsgren, M., Jaki, M., Norin, M., Sterky, C., Enhorning, A. *et al.* (1997). Identification of framework residues in a secreted recombinant antibody fragment that control production level and localization in *Escherichia coli*. *J. Biol. Chem.* **272**, 12430–12436.
11. Kipriyanov, S. M., Moldenhauer, G., Martin, A. C., Kupriyanova, O. A. & Little, M. (1997). Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bacterial secretion but not the affinity. *Protein Eng.* **10**, 445–453.
12. Nieba, L., Honegger, A., Krebber, C. & Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10**, 435–444.
13. Wörn, A. & Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010.
14. Wörn, A. & Plückthun, A. (1999). Different equilibrium stability behavior of ScFv fragments: identification, classification, and improvement by protein engineering. *Biochemistry*, **38**, 8739–8750.
15. Jäger, M. & Plückthun, A. (1999). Domain interactions in antibody Fv and scFv fragments: effects on unfolding kinetics and equilibria. *FEBS Letters*, **462**, 307–312.
16. Jäger, M., Gehrig, P. & Plückthun, A. (2001). The scFv fragment of the antibody hu4D5-8: evidence for early premature domain interaction in refolding. *J. Mol. Biol.* **305**, 1111–1129.
17. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G. *et al.* (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86.
18. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nature Biotechnol.* **18**, 1287–1292.
19. Wirtz, P. & Steipe, B. (1999). Intrabody construction and expression III: engineering hyperstable VH domains. *Protein Sci.* **8**, 2245–2250.
20. Jäger, M. & Plückthun, A. (1999). Folding and assembly of an antibody Fv fragment, a heterodimer stabilized by antigen. *J. Mol. Biol.* **285**, 2005–2019.
21. Ohage, E. & Steipe, B. (1999). Intrabody construction and expression. I. The critical role of VL domain stability. *J. Mol. Biol.* **291**, 1119–1128.
22. Stevens, F. J. (2000). Four structural risk factors identify most fibril-forming kappa light chains. *Amyloid*, **7**, 200–211.
23. Honegger, A. & Plückthun, A. (2001). Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *J. Mol. Biol.* **309**, 657–670.
24. Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B., Henner, D., Wong, W. L. *et al.* (1992). Humanization

of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl Acad. Sci. USA*, **89**, 4285–4289.

25. Ewert, S., Cambillau, C., Conrath, K. & Plückthun, A. (2002). Biophysical properties of camelid VHH domains compared to those of human V$_H$3 domains. *Biochemistry*, **41**, 3628–3636.

26. Power, B. E., Ivancic, N., Harley, V. R., Webster, R. G., Kortt, A. A., Irving, R. A. & Hudson, P. J. (1992). High-level temperature-induced synthesis of an antibody VH-domain in *Escherichia coli* using the PelB secretion signal. *Gene*, **113**, 95–99.

27. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.

28. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure, A Practical Approach* (Creighton, T. E., ed.), pp. 299–321, Oxford University Press, New York.

29. Raffen, R., Dieckman, L. J., Szpunar, M., Wunschl, C., Pokkuluri, P. R., Dave, P. *et al.* (1999). Physico-chemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains. *Protein Sci.* **8**, 509–517.

30. Honegger, A. & Plückthun, A. (2001). The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domains. *J. Mol. Biol.* **309**, 687–699.

31. Nakamura, H. (1996). Roles of electrostatic interaction in proteins. *Quart. Rev. Biophys.* **29**, 1–90.

32. Pace, C. N. (1990). Measuring and increasing protein stability. *Trends Biotechnol.* **8**, 93–98.

33. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOL-MOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. See also pp. 29–32.

34. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (1993). X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969–995.

35. Saul, F. A. & Poljak, R. J. (1993). Structural patterns at residue positions 9, 18, 67 and 82 in the VH framework regions of human and murine immunoglobulins. *J. Mol. Biol.* **230**, 15–20.

36. Jung, S., Spinelli, S., Schimmele, B., Honegger, A., Pugliese, L., Cambillau, C. & Plückthun, A. (2001). The importance of framework residues H6, H7 and H10 in antibody heavy chains: experimental evidence for a new structural subclassification of antibody VH domain. *J. Mol. Biol.* **309**, 701–716.

37. Fink, A. L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **3**, R9–23.

38. Jung, S., Honegger, A. & Plückthun, A. (1999). Selection for improved protein stability by phage display. *J. Mol. Biol.* **294**, 163–180.

39. Ewert, S., Honegger, A. & Plückthun, A. (2002). Structural-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. *Biochemistry*. In press.

40. Schiweck, W., Buxbaum, B., Schatzlein, C., Neiss, H. G. & Skerra, A. (1997). Sequence analysis and bacterial production of the anti-c-myc antibody 9E10: the VH domain has an extended CDR-H3 and exhibits unusual solubility. *FEBS Letters*, **414**, 33–38.

41. van der Linden, R. H., Frenken, L. G., de Geus, B., Harmsen, M. M., Ruuls, R. C., Stok, W. *et al.* (1999). Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochim. Biophys. Acta*, **1431**, 37–46.

42. Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R. *et al.* (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Struct. Biol.* **3**, 803–811.

43. Decannière, K., Desmyter, A., Lauwereys, M., Ghahroudi, M. A., Muyldermans, S. & Wyns, L. (1999). A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Struct. Fold. Des.* **7**, 361–370.

44. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L. *et al.* (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260.

45. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C. *et al.* (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnol.* **14**, 309–314.

46. Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. *J. Mol. Biol.* **227**, 776–798.

47. Glockshuber, R., Schmidt, T. & Plückthun, A. (1992). The disulfide bonds in antibody variable domains: effects on stability, folding *in vitro*, and functional expression in *Escherichia coli*. *Biochemistry*, **31**, 1270–1279.

48. Wörn, A., Auf der Maur, A., Escher, D., Honegger, A., Barberis, A. & Plückthun, A. (2000). Correlation between *in vitro* stability and *in vivo* performance of anti-GCN4 intrabodies as cytoplasmic inhibitors. *J. Biol. Chem.* **275**, 2795–2803.

49. Visintin, M., Settanni, G., Maritan, A., Graziosi, S., Marks, J. D. & Cattaneo, A. (2002). The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. *J. Mol. Biol.* **317**, 73–83.

50. Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, G. T. & Rabbitts, T. H. (2002). Intracellular antibody capture technology: application to selection of intracellular antibodies recognising the BCR-ABL oncogenic protein. *J. Mol. Biol.* **317**, 85–94.

51. Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A. M. *et al.* (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nature Biotechnol.* **18**, 852–856.

52. Pini, A., Viti, F., Santucci, A., Carnemolla, B., Zardi, L., Neri, P. & Neri, D. (1998). Design and use of a phage display library. Human antibodies with sub-nanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. *J. Biol. Chem.* **273**, 21769–21776.

53. Holt, L. J., Bussow, K., Walter, G. & Tomlinson, I. M. (2000). By-passing selection: direct screening for antibody–antigen interactions using protein arrays. *Nucl. Acids Res.* **28**, E72.

54. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R. & Plückthun, A. (1997). Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods*, **201**, 35–55.

55. Bothmann, H. & Plückthun, A. (1998). Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nature Biotechnol.* **16**, 376–380.

56. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*. NIH Publication No. 91-3242, National Technical Information Service (NTIS), Springfield, VA.

57. Bass, S., Gu, Q. & Christen, A. (1996). Multicopy suppressors of prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pA. *J. Bacteriol.*, **178**, 1154–1161.

58. Buchner, J. & Rudolph, R. (1991). Renaturation, purification and characterization of recombinant Fab-fragments produced in *Escherichia coli*. *Biotechnology (N Y)*, **9**, 157–162.

59. Lindner, P., Bauer, K., Krebber, A., Nieba, L., Kremmer, E., Krebber, C. *et al.* (1997). Specific detection of his-tagged proteins with recombinant anti-His tag scFv-phosphatase or scFv-phage fusions. *Biotechniques*, **22**, 140–149.

60. Liu, N., Deillon, C., Klauser, S., Gutte, B. & Thomas, R. M. (1998). Synthesis, physicochemical characterization, and crystallization of a putative retro-coiled coil. *Protein Sci.* **7**, 1214–1220.